THE BULLETIN OF SYMBOLIC LOGIC
Volume 10, Number 2, June 2004

PREDICATIVE FRAGMENTS OF FREGE ARITHMETIC

ØYSTEIN LINNEBO

Abstract. Frege Arithmetic (FA) is the second-order theory whose sole non-logical axiom is Hume's Principle, which says that the number of Fs is identical to the number of Gs if and only if the Fs and the Gs can be one-to-one correlated. According to Frege's Theorem, FA and some natural definitions imply all of second-order Peano Arithmetic. This paper distinguishes two dimensions of impredicativity involved in FA—one having to do with Hume's Principle, the other, with the underlying second-order logic—and investigates how much of Frege's Theorem goes through in various partially predicative fragments of FA. Theorem 1 shows that almost everything goes through, the most important exception being the axiom that every natural number has a successor. Theorem 2 shows that the Successor Axiom cannot be proved in the theories that are predicative in either dimension.

§1. Introduction. Frege is a *logicist* about arithmetic: He holds that pure logic provides a possible source of knowledge of arithmetic and that arithmetic therefore is *a priori*. His defense of this view proceeds in two steps.[1]

First he argues that numbers are ascribed to concepts and that the fundamental law governing such ascriptions is Hume's Principle,[2] which says that the number of Fs is identical to the number of Gs if and only if the Fs and the Gs can be one-to-one correlated. This principle can be formalized as

$$\text{(HP)} \qquad Nx.Fx = Nx.Gx \leftrightarrow F \approx G$$

where 'N' is an operator that applies to concept variables to form singular terms, and where $F \approx G$ is a second-order formula saying that there is a relation R that one-to-one correlates the Fs and the Gs.

Then Frege gives an explicit definition of terms of the form $Nx.Fx$. He does this in a theory of extensions consisting of second-order logic plus an axiom stating that the extension of F is identical to the extension of G if and only if the Fs and the Gs are co-extensional:

$$\text{(V)} \qquad \hat{x}.Fx = \hat{x}.Gx \leftrightarrow \forall x(Fx \leftrightarrow Gx).$$

Received July 7, 2003; revised September 7, 2003.

[1] Frege gives an informal exposition of his view in [18]; for a more formal treatment, see [19].

[2] Although *Cantor's Principle* would have been more accurate historically.

1079-8986/04/1002-0001/$3.20

In this theory, Frege defines $Nx.Fx$ as the extension of the concept: is an extension of some concept equinumerous with F. That is, he defines

$$Nx.Fx := \hat{x}(\exists G(x = \hat{u}Gu \wedge F \approx G)).$$

One easily verifies that this definition satisfies HP. Moreover, Frege shows that, in the presence of this definition, his theory of extensions entails all of ordinary arithmetic.

However, as Russell communicated to Frege in a letter of 1902 [35], Frege's theory of extensions is inconsistent: Russell's paradox is derivable in it. This eventually led Frege to give up on logicism.

Neo-logicists, such as Bob Hale and Crispin Wright, attempt to get around the problem posed by Russell's paradox by eliminating altogether the second stage of Frege's approach, making do with HP instead.[3] Although HP may not be a logical principle, the neo-logicists argue that it is an explanation of the meaning of the N-operator and provides *a priori* knowledge of HP. Neo-logicism is made possible by two relatively recent technical discoveries. The first discovery is that HP, unlike V, is consistent.[4] In fact, George Boolos showed that Frege Arithmetic (FA)—the second-order theory with HP as its sole non-logical axioms—is equiconsistent with second-order Peano Arithmetic.[5] The second discovery is that FA and some very natural definitions suffice to derive all the axioms of second-order Peano Arithmetic.[6] This result is known as *Frege's Theorem*.[7] It is an amazing result. For more than a century now, informal arithmetic has almost without exception been given some Peano-Dedekind style axiomatization. These axiomatizations regard the natural numbers as finite ordinals, individuated by their position in an omega-sequence. Frege's Theorem shows that an alternative and conceptually completely different axiomatization of arithmetic is possible, based on the idea that the natural numbers are finite cardinals, individuated by the cardinalities of the concepts whose numbers they are.

In this paper I distinguish two independent dimensions of impredicativity involved in Frege Arithmetic—one having to do with HP itself, the other, with the background second-order logic (Section 2). Then I prove versions of Frege's Theorem for various partially predicative subsystems of FA (Section 3). Theorem 1 shows that these subsystems entail most of ordinary (Dedekind-Peano style) arithmetic, the most notable exception being the Successor Axiom (SA), which says that every natural number has a successor. Theorem 2 shows that the Successor Axiom cannot be proved in

[3] See [39] and [21], especially essay 12.

[4] This result was proved already in [20], pp. 446–47. Independently of this, it was conjectured in [39] and proved in [10] and [22].

[5] See [6], which also provides a superb presentation and discussion of the consistency result.

[6] This result was hinted at in [32] and explicitly stated and discussed in [39]. For a nice proof, see [7].

[7] See [23] for a powerful argument that Frege himself knew how to prove Frege's Theorem.

the theories that are predicative in either dimension, as these theories have models in which SA is false (Section 4). In addition to being of independent technical interest, the fragments of Frege Arithmetic that are investigated in this paper are important to the philosophical assessment of neo-logicism. In particular, although I am in general no skeptic about impredicative comprehension, I argue that it is philosophically problematic for the neo-logicists to base their proof of SA on such comprehension (Section 5).

§2. Two dimensions of impredicativity. A definition is said to be *impredicative* if it quantifies over a totality to which the referent of the term defined will belong if the definition succeeds.

The first dimension of impredicativity involved in Frege Arithmetic arises as follows. HP and V belong to a broader class of principles of the form

$$(*) \qquad \Sigma(\alpha) = \Sigma(\beta) \leftrightarrow \alpha \sim \beta$$

where α and β are variables, and where $\sim$ is some equivalence relation on the sort of entities that α and β range over. Principles of this form are known as *abstraction principles*. If α and β are object variables, $(*)$ is said to be an *objectual* abstraction principle, and if α and β are concept variables, $(*)$ is a *conceptual* abstraction principle. When $(*)$ is like HP and unlike V in being an acceptable abstraction principle (whatever exactly that involves), the neo-logicists regard $(*)$ as an implicit definition of the operator 'Σ'. Thus the question of impredicativity arises: $(*)$ is *impredicative* if its right-hand side quantifies over the sort of entities that Σ-terms (such as '$\Sigma(\alpha)$' and '$\Sigma(\beta)$') purport to denote; otherwise, it is *predicative*. Since numbers are supposed to be objects, HP is impredicative. (As is V).

However, there are some closely related predicative abstraction principles, namely the *two-sorted* principles where the terms on the left-hand side belong to a separate logical sort from that of the object variables on the right-hand side. This ensures that the objects denoted by the Σ-terms won't be in the range of the quantifiers on the right-hand side of $(*)$. In the next section I will investigate a two-sorted cousin of HP, which will be called *HP2S*.

Predicative abstraction principles are philosophically interesting because the most natural account of such principles turns out to justify only the predicative ones. According to this account, the task of an abstraction principle $(*)$ is to lay down criteria of identity for the objects to which its left-hand side purports to refer: $\Sigma(\alpha)$ is to be identical to $\Sigma(\beta)$ just in case $\alpha \sim \beta$.[8] But when $(*)$ is impredicative, the condition $\alpha \sim \beta$ for the objects $\Sigma(\alpha)$ and $\Sigma(\beta)$ to be identical quantifies over a totality to which these objects themselves will belong if the definition succeeds. But this quantification

[8] I argue in [29] that Frege's own account of abstraction principles was of this kind. Michael Dummett and Crispin Wright discuss the legitimacy of impredicative abstraction principles in [13], [41], and [40]. See also [16], Section II.5.

doesn't make sense unless all objects in its range *already* have criteria of identity. There is no such problem when $(*)$ is predicative. For then, the condition $\alpha \sim \beta$ doesn't involve the criteria of identity we are attempting to lay down. Whether there exist alternative accounts of abstraction principles which justify impredicative principles as well is an exciting philosophical question, which cannot be addressed here.

The second dimension of impredicativity has to do with the background second-order logic. (In order to formalize HP, we need at least *dyadic* second-order logic. In what follows I will always assume full polyadic second-order logic.) When formalizing second-order logic it is useful to restrict the inference rules governing the second-order quantifier such that they can be directly applied only to free variables, not to arbitrary predicative expressions. This restriction forces us to be explicit about what instances of the second-order variables we accept as legitimate. We make this explicit by means of so-called *comprehension axioms*, which are the universal closures of axioms of the form

$$\text{(Comp)} \qquad \exists R \forall u_1 \ldots \forall u_n [R u_1 \ldots u_n \leftrightarrow \phi(u_1, \ldots, u_n)]$$

where R does not appear in ϕ (which is called *the comprehension formula*). Let's consider an example. With the restriction on the rules governing the second-order quantifiers we cannot use UI to infer directly from $\forall F \exists x Fx$ that $\exists x \phi(x)$. Instead, we will have to use the comprehension axiom $\exists X \forall x (Xx \leftrightarrow \phi(x))$ to infer that $\forall x (Xx \leftrightarrow \phi(x))$; then use UI on $\forall F \exists x Fx$ to establish that $\exists x Xx$; and hence finally conclude that $\exists x \phi(x)$.

A comprehension axiom is *predicative* if the comprehension formula ϕ contains no bound second-order variables, and *impredicative* otherwise. This terminology makes sense because impredicative comprehension axioms define second-order entities by quantifying over totalities to which these entities belong. A more fine-grained definition goes as follows. A formula ϕ is said to be Π^1_n, for some $n > 0$, when all of its second-order quantifiers occur as a block at the beginning of the formula, and when this block begins with universal second-order quantifiers and contains $n - 1$ alterations of second-order quantifiers; ϕ is said to be Π^1_0 (or *predicative*) if it contains no second-order quantifiers. ϕ is said to be Σ^1_n if it is the negation of a Π^1_n-formula. ϕ is said to be Δ^1_n (in a theory T) if ϕ is Π^1_n but provably equivalent (in T) to some Σ^1_n-formula. A comprehension axiom is said to be Π^1_n, Σ^1_n, or Δ^1_n when the comprehension formula ϕ is respectively Π^1_n, Σ^1_n, or Δ^1_n. Because ϕ may contain *free* second-order variables, the class of relations definable by Π^1_n comprehension axioms is closed under Π^1_n definitions. This will be important in what follows. In particular, since the complement of a Π^1_n-definable relation R is predicatively definable relative to R, we may henceforth ignore Σ^1_n comprehension axioms.

There are both technical and philosophical reasons to be interested in subsystems of Frege Arithmetic with limited comprehension. The technical interest of such systems lies in the fact that the full impredicative comprehension scheme is logically very strong, which prompts the question what can be done in subsystems where the principle is weakened. And the systems with predicative comprehension are philosophically interesting because impredicative comprehension axioms put greater demands on the interpretation of second-order logic than do predicative ones: there are reasons to think that impredicative comprehension is justified only when the second-order variables range over arbitrary subcollections of the first-order domain. I attempt to draw some consequences of this in Section 5.

The two dimensions of impredicativity involved in Frege Arithmetic are independent of one another, logically as well as conceptually. For our choice whether or not to allow impredicative comprehension is not affected by our choice whether or not to allow impredicative abstraction principles. Indeed, in the next section I describe formal theories corresponding to each of the two binary choices.

§**3. Frege's Theorem.** Frege's Theorem says that, in the presence of some very natural definitions, Frege Arithmetic implies all the axioms of second-order Peano Arithmetic. This section will investigate how much of Peano Arithmetic can be recovered from various subsystems of Frege Arithmetic. Before our results can be stated, however, we need some definitions.

Definition 1 (Neo-logicist theories of arithmetic). Let 'N' be an operator that applies to concept variables to form singular terms, and let Hume's Principle be the axiom

$$\text{(HP)} \qquad Nx.Fx = Nx.Gx \leftrightarrow F \approx G.$$

Let L_{FA} be the second-order language whose only non-logical symbol is the N-operator. Let L_{FA2S} be as L_{FA} except that it is based on two disjoint sorts—one arithmetical and one non-arithmetical—and that the N-operator applies only to non-arithmetical concept variables to form arithmetical singular terms. Let *HP2S* be as HP except that it contains this alternative N-operator.

(a) *Frege Arithmetic*, or FA, is the second-order theory based on L_{FA} with full impredicative comprehension and whose sole non-logical axiom is HP.
(b) *Two-sorted Frege Arithmetic*, or FA2S, is the second-order theory based on L_{FA2S} with full impredicative comprehension on both sorts and whose sole non-logical axiom is HP2S.
(c) Π^1_n-FA is like FA except that only Π^1_n-comprehension is allowed; and likewise for Δ^1_n-FA. In particular, *predicative Frege Arithmetic* is the theory Π^1_0-FA.[9]

[9]It may seem problematic that this theory allows the N-operator to occur in comprehension formulas that are supposed to be predicative. For the implicit definition of this operator

(d) Π^1_n-FA2S is like FA2S except that only Π^1_n-comprehension is allowed in either sort; and likewise for Δ^1_n-FA2S. In particular, *predicative two-sorted Frege Arithmetic* is the theory Π^1_0-FA2S.

REMARK 1. There is also a *schematic* version of HP

$$(\mathrm{HP}^+) \qquad Nx.\phi(x) = Nx.\psi(x) \;\leftrightarrow\; \phi(x) \approx_x \psi(x)$$

where 'N' is a variable-binding operator that applies to *arbitrary formulas* of the language to form singular terms, and where the lower index on $\approx$ specifies the variable with respect to which the one-to-one correlation is said to exist. What is the relation between the axiomatic and the schematic forms of Hume's Principle?

Clearly, the schematic form HP^+ entails the axiomatic form HP. For by instantiating HP^+ with respect to free second-order variables F and G and then universally generalizing, we get HP. Conversely, HP is easily seen to imply all instances of HP^+ whose formulas $\phi(x)$ and $\psi(x)$ are allowed to figure in comprehension axioms. However, when full impredicative comprehension isn't assumed, HP will be weaker than HP^+.

DEFINITION 2. (a) When one of the above neo-logicist theories has comprehension for $\phi(x)$, we can write $Nx.\phi(x)$—although this isn't part of the official syntax—with the convention that any context $\psi(Nx.\phi(x))$ be understood as $\exists F[\psi(Nx.Fx) \wedge \forall x(Fx \leftrightarrow \phi(x))]$.

(b) Whenever a neo-logicist theory T in language L has restricted comprehension, let L^+ be as L except that the N-operator is allowed to apply to arbitrary formulas, and let T^+ be the L^+-theory which is as T except that it has HP^+ as an axiom scheme instead of HP as a single axiom.[10]

DEFINITION 3. Let F be a predicate variable, and let $\rho(x, y)$ be a formula with x and y free.

(i) F is said to be *ρ-hereditary* iff $\forall x \forall y(Fx \wedge \rho(x, y) \rightarrow Fy)$; we abbreviate this as $\mathrm{Her}_\rho(F)$.

(ii) Let $\mathrm{Suc}_\rho(x, F)$ abbreviate the claim $\forall z(\rho(x, z) \rightarrow Fz)$ that all ρ-successors of x are F.

(iii) *The proper ancestral* $\rho^*(x, y)$ of ρ is defined by $\forall F[(\mathrm{Suc}_\rho(x, F) \wedge \mathrm{Her}_\rho(F)) \rightarrow Fy]$.

(iv) *The improper ancestral* of ρ, written $\rho^{*=}(x, y)$, is defined as $\rho^*(x, y) \vee x = y$.

---

relates it to a formula that quantifies over relations. However, the restriction to predicative comprehension is compatible with a stepwise extension of the language in which the comprehension formulas are given as we come to understand new expressions. Now, on the view in question, HP is solely responsible for fixing the meaning of the N-operator. It therefore makes sense to allow the operator to occur in predicative comprehension formulas.

[10] Π^1_0-FA$^+$ will figure prominently in Section 4. Note that when T^+ has a comprehension axiom with comprehension formula ϕ, then ϕ has all its second-order quantifiers in front, never in the scope of an embedded N-term.

REMARK 2. What we have done is, in effect, to define that $\rho^*(x, y)$ is to hold just in case y is such that induction on ρ with basis x can be used to prove results about y. Induction is thus "built into" the definition of the ancestral. Because of our convention regarding the inference rules governing the second-order quantifiers, this means we get only as much induction as we allow through our comprehension axioms.

LEMMA 1 (Frege [17]). *Let $\rho(x, y)$ be as above and $\rho^*(x, y)$ its ancestral. Then we have*:

(a) $\rho(x, y) \to \rho^*(x, y)$,
(b) $\rho^*(x, y) \wedge \rho^*(y, z) \to \rho^*(x, z)$,
(c) $\rho^*(x, z) \to \rho(x, z) \vee \exists y[\rho^*(x, y) \wedge \rho(y, z)]$.

PROOF. The proofs of (a) and (b) are easy. For (c), let F be the concept $\lambda u[\rho(x, u) \vee \exists y(\rho^*(x, y) \wedge \rho(y, u))]$. We prove by induction that F holds of all z such that $\rho^*(x, z)$. The basis case, $\rho(x, y)$, is trivial. Next we need to prove $\mathrm{Her}_\rho(F)$. Assume Fa and $\rho(a, b)$. By the definition of F, either $\rho(x, a)$ or there is some y such that $\rho^*(x, y)$ and $\rho(y, a)$. In the former case, we have $\rho^*(x, a)$ and $\rho(a, b)$, from which it follows that Fb. In the latter case, we have $\rho^*(x, a)$ and $\rho(a, b)$, from which it also follows that Fb. So F is ρ-hereditary. Hence the lemma follows by induction. $\dashv$

Note on the use of comprehension. (a) and (b) require no comprehension. For (c), assume $\rho(x, y)$ is Π^1_n. Then use Π^1_n-comprehension to establish the existence of a dyadic relation R_1 corresponding to $\rho(x, y)$. Then use Π^1_1-comprehension with R_1 as a parameter to establish the existence of a relation R_2 corresponding to $\rho^*(x, y)$. The induction property F is predicatively definable relative to R_1 and R_2. So for (c), Π^1_k-comprehension suffices, where $k = \max\{1, n\}$.

LEMMA 2. *Predicative comprehension suffices to prove*:

$$\rho^*(x, z) \to \exists y\, \rho(y, z).$$

PROOF. Assume $\neg\exists y\, \rho(y, z)$. By predicative comprehension, let F be the property of not being identical to z. Then the successors of x have F, and F is hereditary. Hence $\neg\rho^*(x, z)$. $\dashv$

DEFINITION 4 (the Frege Definitions). In (both one- and two-sorted) Frege Arithmetic we make the *Frege Definitions*:

(DZ) $0_f = Nx.x \neq x$.

(DP) $P(x, y) \leftrightarrow \exists F \exists w[Fw \wedge x = Nu(Fu \wedge u \neq w) \wedge y = Nu.Fu]$.

(DN) $\mathbb{N}(x) \leftrightarrow P^{*=}(0_f, x)$.

In one-sorted Frege Arithmetic, other *Frege numerals* are defined recursively by

$$(n+1)_f = Nx(x=0_f \vee x=1_f \vee \ldots \vee x=n_f).^{11}$$

REMARK 3. For technical purposes, all these definitions do is introduce certain new symbols as abbreviations of certain strings of old symbols. But for philosophical purposes, it may be important that the definitions provide plausible analyses of our pre-theoretic arithmetical notions: that *zero* is the number of non-self-identical things; that x *immediately precedes* y just in case x is the number of some concept under which fall all but one of the objects falling under some concept whose number is y; and that x is a *natural number* just in case it follows zero in the P-series.

DEFINITION 5 (Dedekind-Peano theories of arithmetic). Let L_{PA^2} be the second-order language whose only non-logical symbols are a singular term '0', a predicate '$\mathbb{N}$', and a binary relation symbol 'P'.

(a) *Second-order Peano Arithmetic* or PA^2 is the second-order theory based on L_{PA^2} with full impredicative comprehension and with arithmetical axioms

(PA1) $\mathbb{N}0$
(PA2) $\mathbb{N}x \wedge Pxy \rightarrow \mathbb{N}y$
(PA3) $Pxy \wedge Pxz \rightarrow y = z$
(PA4) $Pxz \wedge Pyz \rightarrow x = y$
(PA5) $\neg\exists x\, Px0$
(PA6) $\forall x(\mathbb{N}x \rightarrow \exists y\, Pxy)$ (the Successor Axiom, or SA)
(PA7) $\forall F[F0 \wedge \mathrm{Her}_P(F) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx)]$

(b) Π^1_n-PA^2 is like PA^2 except that only Π^1_n-comprehension is allowed; and likewise for Δ^1_n-PA^2.

(c) A is like Π^1_0-PA^2 except that comprehension is allowed only on formulas containing no occurrences of '$\mathbb{N}$' or 'P' and that it has the Predecessor Axiom $\forall y(\mathbb{N}y \wedge y \neq 0 \rightarrow \exists x\, Pxy)$.

(d) Whenever T is a subtheory of PA^2, T^- is T minus the Successor Axiom.

REMARK 4. Our axiomatization of Peano Arithmetic is slightly unusual in that it allows for non-numbers in its domain. However, this is the version of Peano Arithmetic that lends itself most readily to comparison with Frege Arithmetic. For Frege Arithmetic would require an additional and completely artificial axiom to rule out non-numbers in its domain.

The predicative subtheories of PA^2 are also unusual in not having primitive signs for addition and multiplication and axioms giving a recursive characterization of these operations. But of course, in systems with the Successor Axiom and at least Π^1_1-comprehension we can give the standard second-order explicit definitions of addition and multiplication.

[11] Recall that, by the convention of Definition 2(a), the expression on the right-hand side may in turn only be contextually defined.

THEOREM 1 (Frege's Theorem). *Relative to the Frege Definitions, we have*:

(a) *FA entails* PA^2.
(b) *FA2S entails* PA^{2-}.
(c) *For subsystems of FA with limited comprehension the following hold*:
   (i) *For all* $n \geq 1$, Π^1_n*-FA entails* Π^1_n*-*PA^2.
   (ii) *For all* $n \geq 2$, Δ^1_n*-FA entails* Δ^1_n*-*PA^2.
   (iii) Π^1_0*-FA entails*
   $A^- \cup \{$*all truths of the forms* $m_f \neq n_f$, $P(m_f, n_f)$, *and* $\neg P(m_f, n_f)\}$.
(d) *For subsystems of FA2S with limited comprehension the following hold*:
   (i) *For all* $n \geq 1$, Π^1_n*-FA2S entails* Π^1_n*-*PA^{2-}.
   (ii) *For all* $n \geq 2$, Δ^1_n*-FA2S entails* Δ^1_n*-*PA^{2-}.
   (iii) Π^1_0*-FA2S entails* A^-.

REMARK 5. (a) is familiar from the works cited in footnote 6. (b) says that, when we reject the first dimension of impredicativity, we lose the Successor Axiom (henceforth only *SA*).[12] The first two clauses of (c) are unsurprising: we get only as much PA^2 comprehension (and thus only as much induction) as we pay for in the form of FA comprehension axioms. But clause (iii) holds two surprises: When impredicative comprehension is given up altogether, it appears we can prove neither SA nor comprehension on predicative formulas of L_{PA^2} containing 'P' and '$\mathbb{N}$'.[13] Finally, (d) is as one would expect, given (b) and (c).

PROOF OF (a). We proceed step by step, listing in square brackets the axioms we have used.

1. To prove: $\mathbb{N}(0)$. Trivial. [DN].
2. To prove: $\mathbb{N}(x) \wedge P(x,y) \to \mathbb{N}(y)$. Trivial. [DN].
3. To prove: $P(x,y) \wedge P(x,z) \to y = z$. Assume $P(x,y)$ and $P(x,z)$. By DP there are properties F and G and objects a and b such that

$$\begin{array}{ll} y = Nu.Fu & z = Nu.Gu \\ \quad Fa & \quad Gb \\ x = Nu(Fu \wedge u \neq a) & x = Nu(Gu \wedge u \neq b) \end{array}$$

   By HP there is a relation R one-to-one correlating $\lambda u(Fu \wedge u \neq a)$ and $\lambda u(Gu \wedge u \neq b)$. Define $R'uv :\leftrightarrow Ruv \vee (u = a \wedge v = b)$ and note that R' one-to-one correlates the Fs and the Gs. By predicative comprehension and HP it follows that $y = z$. [HP, DP, Π^1_0-CA].
4. To prove: $P(x,z) \wedge P(y,z) \to x = y$. Assume $P(x,z)$ and $P(y,z)$. By DP there are properties F and G and objects a and b such that

$$\begin{array}{ll} z = Nu.Fu & z = Nu.Gu \\ \quad Fa & \quad Gb \\ x = Nu(Fu \wedge u \neq a) & y = Nu(Gu \wedge u \neq b) \end{array}$$

[12] This result is established and discussed in [26].

[13] The former surprise is remarked upon in [24], p. 219 and [27], p. 192.

By HP there is a relation R that one-to-one correlates the Fs and the Gs. Using R as parameter, we can predicatively define a relation that one-to-one correlates $\lambda u(Fu \wedge u \neq a)$ and $\lambda u(Gu \wedge u \neq b)$. So by HP we get $x = y$. [HP, DP, Π^1_0-CA].

5. To prove: $\neg\exists x\, P(x, 0)$. If $P(x, 0)$, then by DP there is a concept F and an object a such that $0 = Nu.Fu$ and Fa. By DZ and HP this means that the non-self-identical objects are one-to-one correlated with the Fs, which is impossible. [HP, DZ, DP].
6. To prove: $\forall x(\mathbb{N}(x) \rightarrow \exists y\, P(x, y))$. This case is by far the hardest and has been separated out as Lemma 3 below. Note that this case requires much stronger axioms than the other cases: [HP, DZ, DP, DN, Π^1_1-CA].
7. To prove: $\forall F[F0 \wedge \mathrm{Her}_P(F) \rightarrow \forall x(\mathbb{N}(x) \rightarrow Fx)]$. Trivial. [DN].

Finally, we must verify that every comprehension axiom of PA^2 follows from some comprehension axiom of FA. So let $\exists R \forall u_1 \dots \forall u_n[Ru_1 \dots u_n \leftrightarrow \phi(u_1, \dots , u_n)]$ be some comprehension axiom of PA^2. The only difficulty is that ϕ may contain occurrences of 'P' and '$\mathbb{N}$', whose definitions contain second-order quantifiers. However, using suitable comprehension axioms, we prove the existence of a relation and a concept corresponding to these two definitions. This allows us to replace the two definitions with suitable free second-order variables. The resulting formula will then be a comprehension axiom of FA. ⊣

Lemma 3. *Relative to the Frege Definitions, Π^1_1-FA entails SA.*

Proof. Let $x < y$ and $x \leq y$ abbreviate $P^*(x, y)$ and $P^{*=}(x, y)$ respectively. We will use induction to show that the following hold for all natural numbers n:

(i) $\neg n < n$,
(ii) $P(n, Nu(u \leq n))$.

From this the Lemma will follow trivially.[14]

Note on the use of comprehension. Some care is needed to ascertain that the induction needs no more than Π^1_1-comprehension; for when defined notions are spelled out directly, $\neg\, n < n$ is Σ^1_2, and $P(n, Nu(u \leq n))$ is Σ^1_3. The trick is iterated applications of Π^1_1-comprehension with carefully chosen parameters. First use Π^1_1-comprehension to establish the existence of a dyadic relation R_1 corresponding to the definition of P. Then use Π^1_1-comprehension with parameter R_1 to establish the existence of a relation R_2 corresponding to $u \leq n$. Relative to R_1 and R_2, Π^1_1-comprehension ensures the existence of a concept corresponding to $P(n, Nu(u \leq n))$. These are the only impredicative comprehension axioms needed in the argument.[15]

---

[14]The idea of proving SA by means of this simultaneous induction is due to Warren Goldfarb.

[15]Working relative to a theory of finite sets, [14] and [15] show how the relations R_1 and R_2 can be predicatively defined, which allows the present induction to go through with only

Now for the induction proper. We begin with the base case. Assume $x < 0$. PA4 allows us to apply Lemma 1 to show that there is some y such that $P(y, 0)$. But this is impossible. So $\neg\exists x(x < 0)$, which establishes (i). It also follows that $u \leq 0 \leftrightarrow u = 0$, which establishes (ii).

For the induction step, assume that (i) and (ii) hold for m and that $P(m, n)$. Assume for contradiction that $n < n$. By Lemma 1, either $P(n, n)$, or there is some y such that $n < y$ and $P(y, n)$. In the former case, it follows from PA4 that $m = n$. But then we have $P(m, m)$, contradicting the induction hypothesis. In the latter case, PA4 yields $y = m$. But then we have $P(m, n)$ and $n < m$, which yields $P(m, m)$ and again a contradiction. Hence $\neg n < n$. Observe that since $\neg n < n$, $Nu(u < n)$ bears P to $Nu(u \leq n)$. So to reach our goal of establishing that n bears P to $Nu(u \leq n)$, it suffices to establish $n = Nu(u < n)$. By the induction hypothesis, we know that m bears P to $Nu(u \leq m)$. Since $P(m, n)$, PA3 yields $n = Nu(u \leq m)$. So all that remains is to show $Nu(u \leq m) = Nu(u < n)$. Now, by Lemma 1 we easily show that $\forall u(u \leq m \leftrightarrow u < n)$. Applying HP to this, we are done. ⊣

Proof of Theorem 1 (b). By inspection of the proof of Theorem 1(a) one easily verifies that, with one exception, the entire proof goes through with HP2S instead of HP. The exception is SA, the proof of which turns on the idea of counting numbers, which isn't allowed in the two-sorted system FA2S. (In fact, in a model of FA2S there need be no numbers larger than the cardinality of the domain of non-numbers. Without any extra-logical information about the number of non-numbers, this means that FA2S cannot prove the existence of numbers larger than 1.) ⊣

Proof of Theorem 1 (c). (i) PA1 through PA7 can be established exactly as in the proof of Theorem 1(a). The same goes for the PA^2 comprehension axioms, once we observe that Π^1_1-comprehension suffices to prove the existence of a relation and a concept corresponding to the definitions of 'P' and '$\mathbb{N}$'. (ii) follows directly from the proof of (i). Concerning (iii), observe that without Π^1_1-comprehension, our proof of SA fails, as does the above procedure for eliminating occurrences of 'P' and '$\mathbb{N}$' from PA^2 comprehension axioms. However, by induction in the meta-language we easily show that Π^1_0-FA proves the relevant truths about individual numbers. The Predecessor Axiom is immediate from Lemma 2. ⊣

Proof of Theorem 1 (d). Straightforward by combining the ideas of the previous two proofs. ⊣

§4. Models in which the successor axiom is false.

The proof of Theorem 1 leaves little doubt that the current proof strategy has been exploited for all it is worth. But this is compatible with the existence of alternative, less obvious

---

predicative comprehension. However, the argument I develop in Section 5 can be generalized to apply to this system as well.

proof strategies which establish stronger results. However, our next theorem shows that in important respects Theorem 1 is optimal.

THEOREM 2. *Relative to the Frege Definitions, we have*:

(a) *FA2S does not entail SA*.

(b) Π^1_0-*FA*$^+$ *does not entail SA* (*thus, nor does the weaker theory* Π^1_0-*FA*).

Theorem 2 is proved by constructing models of FA2S and Π^1_0-FA$^+$ in which SA is false. For FA2S, this is trivial, as indicated in the proof of Theorem 1(b). But for Π^1_0-FA$^+$, the construction of a suitable model is somewhat tricky and will take up the remainder of this section.

What would a model of Π^1_0-FA$^+$ in which SA fails look like? Obviously, there would have to be an x such that $P^*(0_f, x)$ and $\neg\exists y\, P(x, y)$. But there are more interesting things to say:

1. x must be a number; that is, there must be some F such that $x = Nu.Fu$. For applying Lemma 2 to $P^*(0_f, x)$, it follows that x has a P-predecessor, which by DP only numbers can have.
2. F must be the universal concept, that is the concept V such that $\forall u.Vu$. For otherwise, we can pick some a such that $\neg Fa$, in which case x would be succeeded by $Nu(Fu \vee u = a)$. Let '$\hat{0}_f$' abbreviate '$Nu.Vu$', and following Boolos [8], let's call its reference *anti-zero*.[16]
3. The axiom $\exists u.\neg Fu \to F \not\approx V$ must hold. I will call this *the Euclidean Axiom* because it resembles the old Euclidean dictum that a whole is always greater than its parts. For assume the Euclidean Axiom fails and that we have $\neg Fa$ and a relation R that one-to-one correlates the Fs with the universe. Then $\hat{0}_f$ is identical to $Nu.Fu$, although this latter is succeeded by $Nu(Fu \vee u = a)$. Conversely, assume $\hat{0}_f$ is succeeded by $Nu.Fu$. Then the universe is one-to-one correlated with all the Fs *except one*, which contradicts the Euclidean Axiom.

[16]In a free logic formalization of Frege Arithmetic, $\exists x(x = \hat{0}_f)$ need not be a theorem. Might a predicative proof of SA be possible in some such system? Not in the two free logic formalization of Frege Arithmetic of which I am aware, [34] and [37], since both have models in which $\exists x(x = \hat{0}_f)$. This is easily verified for the theory of [37], since all its axioms concerning the existence of numbers are positive (saying that certain numbers exist if certain conditions are met), never negative (denying the existence of certain numbers). The theory of [34], however, says that there is such an object as the number of Fs just in case F is either empty or is equinumerous with some bounded initial segment of a well-ordered domain. But in fact, all set size models of this theory can be extended so that V satisfies this condition: Choose some R that well-orders the first-order domain and make sure that the second-order domain contains the relation R' which is like R except that the R-initial element is R'-final.

Of course, we could explicitly deny the existence of $\hat{0}_f$ by adopting the axiom $\exists x(x = Nu.Fu) \leftrightarrow \neg\forall uFu$. However, this axiom would not only be philosophically unmotivated, but it would be true in a straightforward modification of the model of Π^1_0-FA$^+$ + $\neg$SA that I am about to construct. (In this modified model, $Nu(u \neq 0_f)$ would be a natural number without a successor.)

4. Recall that Π^1_0-FA proves $m_f \neq n_f$ whenever $m \neq n$. Anti-numerals other than '$\hat{0}_f$' can be introduced by letting $\ulcorner \hat{n}_f \urcorner$ abbreviate $\ulcorner Nu(u \neq 1_f \wedge \ldots \wedge u \neq n_f)\urcorner$. Π^1_0-FA plus the Euclidean Axiom proves $\hat{m}_f \neq \hat{n}_f$ whenever $m \neq n$, as well as $m_f \neq \hat{n}_f$ for any m and n. A model of $\neg$SA must therefore assign distinct objects to all the Frege numerals and anti-numerals.

The following lemma sums up these results.

LEMMA 4. *$\mathcal{M} \models \Pi^1_0$-$FA^+ + \neg SA$ iff $\mathcal{M} \models \Pi^1_0$-$FA^+ + P^*(0_f, \hat{0}_f)$ + the Euclidean Axiom.*

Before we begin our construction of a model of Π^1_0-FA$^+ + \neg$SA, we need three lemmas.

LEMMA 5. *Assume the dyadic relation R on a domain D is definable by a quantifier-free formula in a language whose only predicate is identity and whose only singular terms are variables. Assume R one-to-one correlates the elements of two subsets A and B of D. Then A and B have the same elements apart from at most finitely many exceptions.*

PROOF. Assume R is defined by the quantifier-free formula $\rho(u, v)$ under a certain assignment to its free variables other than u and v. Assume, for contradiction, that the Lemma is false. Then we can find two elements x and y of D which are different from each other and from all elements assigned to the free variables of ρ, such that $\langle x, y\rangle$ satisfies $\rho(u, v)$. But since x and y differ from all the elements assigned to ρ's free variables, ρ cannot distinguish between $\langle x, y\rangle$ and any other pair $\langle x', y'\rangle$ of such objects. But then R cannot one-to-one correlate the elements of A and B. ⊣

LEMMA 6. *Let $\phi(u_1, \ldots, u_n)$ be a first-order formula whose only predicate is identity and all of whose singular terms are the free variables shown. Then there is a formula ψ such that*

(a) Π^1_0-$FA \vdash \phi(u_1, \ldots, u_n) \leftrightarrow \psi(u_1, \ldots, u_n)$

(b) *ψ is a disjunction of conjunctions, each conjunction containing, for each identity statement involving the u_i, either it or its negation, but no other conjuncts*

PROOF. By the Löwenheim-Behmann theorem[17] there is a formula θ, provably equivalent (in pure first-order logic) to ϕ, which fits the above description of ψ except that it may also contain some additional conjuncts concerning the number of v distinct from all the u_i. These additional conjuncts say either that there are exactly k such v for some $k < n$, or that there are at least n such v. But by Theorem 1(c)(iii), Π^1_0-FA proves the existence of infinitely many objects. So any one of the additional conjuncts can either be proved, in which case the additional conjunct itself can be deleted, or disproved, in which case the whole conjunction of which the additional conjunct is part can be deleted. This yields the desired ψ. ⊣

[17] See [1] and [31].

LEMMA 7. *Let* $t = \ulcorner Nu.\phi(u, v_1, \ldots, v_n)\urcorner$ *be an* N*-term of* L^{+}_{FA} *with no second-order variables and all of whose free first-order variables are shown. Then* Π^1_0*-*FA^+ *proves a conjunction of conditionals, the antecedents of which consist of statements of identity and non-identity flanked by the* v_i *and by Frege numerals and anti-numerals, and the consequents of which consist of an identity statement flanked by* t *and some particular Frege numeral or anti-numeral. In particular, if* t *is closed, we can prove an identity statement relating* t *to some Frege numeral or anti-numeral.*

PROOF. Let the *rank* of an N-term t be 0 if t is a Frege numeral or anti-numeral; otherwise, let the rank of t be $n + 1$, where n is the highest rank of any embedded N-term. Since the rank of any N-term is finite, it suffices to show that the Lemma holds for N-terms of rank 1; for this allows us to reduce by 1 the rank of any given N-term.

So consider an N-term t of rank 1. By the definition of rank, t contains no N-terms other than Frege numerals and anti-numerals. So we may assume t is $Nu.\phi(u, v_1, \ldots, v_m, a_1, \ldots, a_n)$, where all free variables and constants are shown, and where all the constants a_j are Frege numerals or anti-numerals. Temporarily replacing the a_j with variables enables us to apply Lemma 6 to show that ϕ can be chosen to be quantifier-free. This allows us to reason semantically about theoremhood. So ϕ may be written as a conjunction of conditionals, each antecedent of which specifies a complete set of relations of identity and non-identity between the v_i and the Frege numerals and anti-numerals, and each consequent of which consists of conditions on u. In fact, one easily sees that the consequents can be chosen either as a disjunction of identities $u = v_i$ and $u = a_j$, or as a conjunction of corresponding non-identities. Then the Lemma follows by HP and truth-functional logic. $\dashv$

PROOF OF THEOREM 2 (b).[18] Let the first-order domain D_1 of our model $\mathcal{M}$ be the set of natural numbers. For every number n, introduce into the language a canonical numeral **n** denoting it. (These numerals must not be confused with the Frege numerals and anti-numerals.) Clearly, if the theory in this extended language is consistent, then so is the original theory. These additional numerals allow us to eliminate free first-order variables, because the assignment of numbers to such variables can be imitated by substituting appropriate canonical numerals for the variables. Let the domain D_2 of relations be the set of relations on $\mathbb{N}$ that are definable by a quantifier-free formula in the language whose only predicate is identity and whose only constants are the canonical numerals. Given an assignment of relations to free second-order variables, we can substitute for any free occurrence of a second-order variable the sentential formula defining the relation assigned to it. This allows us to eliminate all free second-order variables.

[18]The organization of this proof draws on that of Heck's proof in [24] of the consistency of the schematic version of Frege's Basic Law V in the context of predicative second-order logic.

We now make assignments to all N-terms of our expanded language containing no bound second-order variables. It suffices to consider closed N-terms, since free variables can be eliminated as described above. Begin by assigning to the Frege numerals m_f and anti-numerals $\hat{n}_f$ the numbers $4m + 1$ and $4n + 3$ respectively. These assignments are in accordance with HP^+: For by Lemma 5, $\mathcal{M}$ satisfies the Euclidean Axiom, and thus by observation 4 at the beginning of this section, we can prove any non-identity flanked by distinct Frege numerals and anti-numerals. Consider now an arbitrary closed N-term $t = \ulcorner Nu.\phi(u)\urcorner$ of the kind in question. We may apply Lemma 7 to t by regarding all of its canonical numerals as variables the assignment to which is held fixed. The resulting N-term reduces to some Frege numeral or anti-numeral once the above assignments to such numerals are taken into account. Interpret t accordingly.

Next we show that Π^1_0-comprehension holds. Consider a predicative comprehension formula ϕ. First eliminate all of its free first- and second-order variables. Then use Lemma 7 to write ϕ as a quantifier-free formula all of whose N-terms are Frege numerals and anti-numerals. Replace these Frege numerals and anti-numerals with canonical numerals in accordance with the assignment made above. This shows that the relation it defines is in D_2.

Finally we interpret all N-terms containing bound second-order variables. As above, it suffices to consider closed N-terms. Let the *degree* of a closed N-term $Nu.\phi(u)$ be 0 if $\phi(u)$ contains no bound second-order variables; otherwise, let its degree be one greater than the greatest degree of any N-term contained in it. Order these N-terms lexicographically in an $\omega \times \omega$ sequence, the first coordinate representing the degree of the N-term, the second, some arbitrary order within this degree. We will make assignments to these N-terms by induction on this sequence, making sure throughout to observe HP^+. For the base case, observe that appropriate assignments have already been made to the N-terms of degree 0. For the induction step, consider an N-term $t = \ulcorner Nu.\phi(u)\urcorner$ of degree m and position n within this degree, and assume appropriate assignments have been made to all earlier N-terms. This allows us to determine the set of numbers satisfying $\phi(u)$, which I will denote by $[\phi(u)]$. Let $J(m,n)$ be some one-to-one function from $\omega \times \omega$ into the even numbers (the odd numbers having been used to interpret the Frege numerals and anti-numerals). If D_2 contains a dyadic relation R that one-to-one correlates $[\phi(u)]$ with some $[\psi(v)]$ where $Nv.\psi(v)$ is an N-term preceding t, assign to t the same number as was assigned to $Nv.\psi(v)$. If there is no such R, assign to t the number $J(m,n)$. These assignments are all in accordance with HP^+.

I claim that the model $\mathcal{M}$ that results from the above construction is as desired. We've seen that $\mathcal{M}$ satisfies Π^1_0-comprehension and that $\mathcal{M}$ was constructed so as to validate HP^+. We've also seen that $\mathcal{M}$ satisfies the Euclidean Axiom. So by Lemma 4 it remains only to show that $\mathcal{M}$ satisfies

$P^*(0_f, \hat{0}_f)$. Assume the set $S \in D_2$ contains the number assigned to 0_f and is closed under the relation assigned to P. Clearly, S must be infinite. Hence it must be cofinite. Hence it must contain some of the numbers assigned to the anti-numerals. Hence, by the closure of S under the relation assigned to P, S must contain the number assigned to $\hat{0}_f$. This means that $\mathcal{M}$ satisfies $P^*(0_f, \hat{0}_f)$. $\dashv$

REMARK 6. It remains unknown whether SA can be derived from either Δ^1_1-FA or Frege Arithmetic with *ramified* second-order logic, using the Frege Definitions.

§5. Philosophical discussion. In this final section I develop two arguments that the neo-logicists' official proof of SA is philosophically problematic. I also observe that Theorem 2(b) limits the neo-logicists' ability to respond to these arguments. Towards the end I discuss some alternative proofs of SA.

The first argument is concerned with the alleged philosophical significance of Frege's Theorem, namely establishing that arithmetical knowledge is *a priori*. For Frege's Theorem to have this philosophical significance, I claim that its proof must have at least a reasonable claim to being just an explication of our ordinary arithmetical reasoning. When the great philosophers down the centuries have pondered the philosophical status of various fields of knowledge, their primary concern has almost always been our *actual* knowledge of the field in question, not how one *might have* come to know these things. For instance, it would have been of little use for Plato to argue that there might have been immortal souls whose knowledge of mathematics was based on pre-natal interactions with immaterial Ideas, or for Kant to argue that there might have been minds equipped with necessary forms giving rise to synthetic *a priori* knowledge of mathematics. But more importantly, faithfulness to ordinary arithmetical reasoning matters because our primary interest is whether *our* knowledge of arithmetic is *a priori*, whether *our* arithmetical thoughts and utterances succeed in referring to abstract objects such as the numbers, and how *these very thoughts and utterances* can be so useful in our dealings with the physical world. To say this is not to deny that it can sometimes be interesting to ask hypothetical questions about how intelligent beings could come to possess knowledge similar to ours and what the philosophical status of this knowledge would be. But even so, this inquiry would supplement, not supplant, the corresponding inquiry about our actual knowledge.[19]

Can the official proof of Frege's Theorem be said to be an explication of our ordinary arithmetical reasoning? As anyone who has reflected on the proof will have noticed, the case of SA sticks out like a sore thumb. The other axioms of PA^2 are proved in very simple and perfectly natural ways.

[19]Similar arguments are developed in [25] and [27].

Arguably, these proofs are *potentially obvious* in the sense that they could be reconstructed by someone with a basic understanding of arithmetic but no mathematical sophistication, given only some very limited prompting.[20] In stark contrast to this, the official proof of SA is anything but obvious. To understand this proof, even mathematical sophisticates need to engage in some pretty serious thought. And to discover the proof in the first place, it took no less of a logician than Frege.

The fact that predicative comprehension is insufficient to prove SA limits the neo-logicists' ability to respond to this argument. For unless they add more axioms or modify the Frege Definitions, any alternative proof of SA will still depend on impredicative comprehension, or at least on ramified second-order logic; and in either case, it seems doubtful that the proof will qualify as an explication of ordinary arithmetical reasoning.

My second argument is based on the suspicion that it is deeply unnatural to base such an elementary arithmetical truth as SA on such strong theoretical resources as impredicative comprehension. The argument is free of any general skepticism about impredicative comprehension and turns on features specific to the neo-logicist project: It aims to show that an adequate understanding of impredicative comprehension requires theoretical resources just as strong as SA itself. The argument proceeds in three steps.

First, I claim that the impredicative comprehension scheme is justified only when the monadic second-order variables are understood as ranging over *arbitrary subcollections* of the first-order domain.[21] (The case of polyadic variables is analogous and will therefore be ignored.) A substantive defense of this view is beyond the scope of the present paper. So I will here limit myself to explaining why the view is plausible.[22] A minimal requirement for definitions by impredicative comprehension to be justified is that we have a conception of a determinate range of possible values of the second-order variables. This requirement is clearly satisfied when the second-order variables are taken to range over arbitrary subcollections of the first-order domain. Moreover, since this range consists of all arbitrary subcollections of the first-order domain, it will be closed under definition by quantification over this range (or under any other mode of definition, for that matter). If, on the other hand, the range of the second-order variables does not contain all arbitrary subcollections of the first-order domain, we will have no guarantee that the range is closed under definition by quantification over this range. Admittedly, we know from Henkin's completeness proof that

[20] For a defense of a closely related claim, see [27].

[21] A note on terminology. I will use the word 'collection' to be neutral between sets and other "set-like" entities, such as mereological sums or pluralities. (For the proposal that second-order logic be interpreted in terms of mereology, see [28]. For the analogous proposal concerning plural quantification, see [4] and [5].) What matters is that the second-order variables range over a determinate totality isomorphic to the powerset of the domain.

[22] See also [33], essays 3 and 8, and [30], Section III.

there are non-standard models of full impredicative second-order logic where the second-order variables do not range over all arbitrary subcollections of the first-order domain. But these models are rather artificial and do not make available any alternative *general* conception of a range of values of the second-order variables.

Second, I claim that our understanding of the notion of an arbitrary subcollection is based on the combinatorial idea of running through the domain, making an independent choice about each element whether or not it is to be included in the subcollection being defined. This claim was first articulated and defended in a classic paper by Bernays [3]. To see why the claim is plausible, it is useful to imagine that you are explaining the notion of an arbitrary subcollection to someone entirely innocent of the notion. It won't do to explain to your pupil that he is to divide the collection in two. For this will either fall short of the idea of an arbitrary subcollection or presuppose it. Rather, you will need to explain to your pupil that, given any collection, he is to make a series of steps, each involving the consideration of one element of the collection and a decision whether or not to include this element in the subcollection.

Third, I claim that the combinatorial idea of running through a domain step by step presupposes an ordinal counterpart of the Successor Axiom. Again, it is useful to imagine you're explaining the idea in question to someone entirely innocent of it. To explain the idea of running through a domain D, deciding which elements to include in a subcollection C, you will have to teach your pupil to carry out the following simple algorithm:

```
WHILE D is non-empty DO
  BEGIN
    pick an element a from D;
    D := D − {a};
    decide whether or not to include a in C
  END
```

Now, for your pupil to understand the WHILE-loop, he needs to understand that no matter how many steps he has carried out, there will always be another step which may need to be carried out. Since the steps in question are abstract possibilities of concrete steps, they are plausibly regarded as ordinals. The principle that for any step there is another is therefore an ordinal counterpart of the Successor Axiom. This means that the neo-logicist proof of SA uses theoretical resources which involve the Ordinal Successor Axiom. The proof therefore threatens to be circular.

Fortunately, there is no need to settle whether the Ordinal Successor Axiom is sufficiently close to SA to make the threatening circle real. It suffices to observe that the theoretical resources involved in the explanation of the notion of an arbitrary subcollection render redundant what appeared to be the greatest virtue of the proof of SA, namely that it gives a logical proof of the existence of infinitely many objects from premises that arguably are analytic. For if our pupil knows the Ordinal Successor Axiom, he will have an infinity of objects at his disposal even before the N-operator is introduced.

If this second argument is sound, Theorem 2(b) shows that the neo-logicists will have to add new axioms or modify the Frege Definitions in order to give a philosophically defensible proof of SA. One such option is to concede that a prior theory of ordinals is needed. This theory of ordinals could then be used to simplify the neo-logicist development of cardinal arithmetic: Rather than rely on the official neo-logicist proof, SA could be established by counting ordinals. However, I think the neo-logicists should be extremely wary of going down this road. If they grant that the theory of ordinals is prior to their own theory of cardinals, it would be easier and much more natural to regard the natural numbers as finite *ordinals* rather than as finite *cardinals*. SA would then just be a special case of the more general Ordinal Successor Axiom. But this would be to give up on neo-logicism, whose guiding idea is that the natural numbers are finite cardinals individuated by HP. In particular, HP would then no longer be a conceptual truth but a substantive claim asserting that the ordinals can be used to measure cardinality, or, more precisely, that when the ordinals are used to tag the elements of a collection, we reach the same tag when counting up the Fs as when counting up the Gs just in case the Fs and the Gs can be one-to-one correlated. When restricted to natural numbers, this claim is indeed a theorem of ordinal arithmetic.[23]

A second option is suggested by recent work of John Burgess [11]. By means of a clever non-standard definition of the natural number predicate, Burgess shows that Robinson Arithmetic Q can be interpreted in Π^1_0-FA, and thus in particular that (a version of) the Successor Axiom can be proved. Burgess keeps the standard Fregean definitions of 0 and the predecessor relation P but defines a non-standard ordering $x \lhd y$ as

$$\exists F \exists G (x = Nu.Fu \wedge F \subset G \wedge y = Nu.Gu)$$

where $F \subset G$ formalizes the claim that the Fs form a proper subcollection of the Gs. Then he defines that x is a *protonatural* if and only if

$$\exists F (\forall y (Fy \leftrightarrow y \lhd x) \wedge x = Nu.Fu \wedge \neg Fx)$$

and proves in Π^1_0-FA that every protonatural has a successor. Perhaps the neo-logicists can use Burgess's technical result to escape my second argument.

An initial difficulty with this suggestion is that mathematical induction doesn't hold of the protonaturals; for instance, the number of numbers of

[23] For the harder direction, from right to left, assume two finite collections can be one-to-one correlated. Then each collection can be one-to-one correlated with an initial segment of the natural numbers. But since the collections themselves can be one-to-one correlated, so can these initial segments. Hence they must be identical.

finite concepts is a protonatural. Since induction is analytic of the concept of natural number if anything is, this prevents the neo-logicists from identifying the concept of natural number with that of a protonatural. To get induction, we need to restrict our attention to those protonaturals x which belong to every concept F to which 0_f belongs and which is closed, *on the protonaturals*, under P. Call such an x a *natural*. Then induction holds on the naturals. Moreover, we can still give a predicative proof of (a version of) the Successor Axiom. For if x is a natural other than 0, Lemma 2 shows that x is a protonatural, whence by Burgess's result it has a successor. So technically speaking, this strategy works. However, a modified version of my first argument applies to the strategy with increased force, as it would be exceedingly implausible to claim that anything like the above notion of a natural is involved in our ordinary arithmetical knowledge. (Of course, Burgess makes no such claim.)

A third option for the neo-logicists is to invoke the modal principle that for any collection of objects it is possible that there be one more object. This principle holds when the objects quantified over are "independent existences," that is, when any one of them could have existed regardless of the existence of the others. And arguably, it is an *a priori* truth that ordinary concrete objects are independent existences in this sense. By adopting this principle as an additional axiom, I believe it is possible to give a neo-logicist proof of SA which is predicative in both dimensions and as natural as the proofs of the other axioms.[24] So this option seems to be the most promising one.

Where does this leave our philosophical discussion? I began this section by arguing that the official neo-logicist proof of SA is philosophically problematic because it is unnatural and because it depends on impredicative comprehension. But if there is another logicist proof of SA which is both natural and fully predicative, the original problem appears to be solved. Even so, however, worries will remain that Frege Arithmetic depends on impredicative comprehension in implausible and problematic ways. Recall from Theorem 1(b) that restricting oneself to predicative comprehension blocks not only the official proof of SA but also the proof of all PA^2 induction axioms whose induction formulas contain occurrences of 'P' and '$\mathbb{N}$'. This is particularly problematic in the case of 'P', as it is hard to see why PA^2 induction axioms talking about such a basic arithmetical relation as succession should depend on impredicative comprehension. The same problem arises for other basic arithmetical relations such as $\textsc{Sum}(m, n, k)$ and $\textsc{Mult}(m, n, k)$, representing addition and multiplication. For in neo-logicism these relations are most naturally defined in second-order terms; for instance, $\textsc{Sum}(m, n, k)$ says that k is the number of some concept equinumerous with the disjunction

[24] See [42] for proof of the Successor Axiom based on this modal principle but motivated by very different concerns. I have developed another such proof, which I argue has the virtues just mentioned.

of two disjoint concepts whose numbers are m and n. In contrast, on the alternative view that regards the natural numbers as ordinals, the successor relation is primitive, and addition and multiplication are predicatively defined in terms of it by the standard recursion axioms. So on this view we trivially get induction on formulas talking about succession, addition, and multiplication without invoking impredicative comprehension. Thus, although a logicist proof of SA which is both natural and predicative will constitute progress, neo-logicism will still be under pressure from non-logicist views which regard the natural numbers as finite ordinals.[25]

REFERENCES

[1] HEINRICH BEHMANN, *Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem*, ***Mathematische Annalen***, vol. 86 (1922), pp. 419–432.

[2] Paul Benacerraf and Hilary Putnam (editors), *Philosophy of mathematics: Selected readings*, Cambridge, Cambridge University Press, 1983, second edition.

[3] PAUL BERNAYS, *On platonism in mathematics*, 1935, reprinted in [2].

[4] GEORGE BOOLOS, *To be is to be a value of a variable (or to be some values of some variables)*, ***Journal of Philosophy***, vol. 81 (1984), pp. 430–449, reprinted in [9].

[5] ———, *Nominalist platonism*, ***Philosophical Review***, vol. 94 (1985), pp. 327–344, reprinted in [9].

[6] ———, *The consistency of Frege's foundations of arithmetic*, ***On beings and sayings: Essays in honor of Richard Cartwright*** (J.J. Thomson, editor), MIT Press, Cambridge, MA, 1987, reprinted in [9] and [12].

[7] ———, *The standard of equality of numbers*, ***Meaning and method: Essays in honor of Hilary Putnam*** (George Boolos, editor), Harvard University Press, Cambridge, MA, 1990, reprinted in [9] and [12].

[8] ———, *Is Hume's principle analytic?*, ***Logic, language, and thought*** (Richard Heck, editor), Oxford University Press, Oxford, 1997, reprinted in [9].

[9] ———, ***Logic, logic, and logic***, Harvard University Press, Cambridge, MA, 1998.

[10] JOHN P. BURGESS, *Review of Crispin Wright's: Frege's conception of numbers as objects*, ***Philosophical Review***, vol. 93 (1984), pp. 638–40.

[11] ———, ***Fixing Frege***, Princeton University Press, Princeton, NJ, forthcoming.

[12] William Demopoulos (editor), ***Frege's philosophy of mathematics***, Harvard University Press, Cambridge, MA, 1995.

[13] MICHAEL DUMMETT, *Neo-Fregeans in bad company?*, 1998, in [36].

[14] SOLOMON FEFERMAN and GEOFFREY HELLMAN, *Predicative foundations of arithmetic*, ***Journal of Philosophical Logic***, vol. 24 (1995), pp. 1–17.

[15] ———, *Challenges to predicative foundations of arithmetic*, ***Between logic and intuition*** (Gila Sher and Richard Tieszen, editors), Cambridge University Press, Cambridge, 2000.

[16] KIT FINE, ***The limits of abstraction***, Oxford University Press, Oxford, 2002.

[17] GOTTLOB FREGE, Begriffsschrift*, a formula language, modeled upon that of arithmetic, for pure thought*, 1879, reprinted in [38].

[25]Thanks to Solomon Feferman, Fernando Ferreira, John MacFarlane, Charles Parsons, and especially John Burgess for useful discussion and comments on earlier versions of this paper. Thanks also to audiences at UC Berkeley, the University of Oslo, and the 12th International Congress of Logic, Methodology, and Philosophy of Science in Oviedo, Spain, where ealier versions were presented.

[18] ———, ***Foundations of arithmetic***, Blackwell, Oxford, 1953, translated by J.L. Austin. Excerpts reprinted in [2].

[19] ———, ***Basic laws of arithmetic***, University of California Press, Berkeley and Los Angeles, 1964, edited and translated by Montgomery Furth.

[20] PETER GEACH, *Review of M. Dummett, Frege: Philosophy of language*, ***Mind***, vol. 84 (1975), pp. 436–499.

[21] BOB HALE and CRISPIN WRIGHT, ***Reason's proper study***, Clarendon, Oxford, 2001.

[22] A.P. HAZEN, *Review of Crispin Wright's: Frege's concept of numbers as objects*, ***Australasian Journal of Philosophy***, vol. 63 (1985), pp. 251–254.

[23] RICHARD G. HECK JR., *The development of arithmetic in Frege's* grundgesetze der arithmetik, ***The Journal of Symbolic Logic***, vol. 58 (1993), pp. 579–601, reprinted in [12].

[24] ———, *The consistency of predicative fragments of Frege's* grundgesetze der arithmetik, ***History and Philosophy of Logic***, vol. 17 (1996), pp. 209–220.

[25] ———, *Finitude and Hume's principle*, ***Journal of Philosophical Logic***, vol. 26 (1997), pp. 598–617.

[26] ———, *The Julius Caesar objection*, ***Language, thought, and logic: Essays in honour of M. Dummett*** (Richard G. Heck Jr., editor), Oxford University Press, Oxford, 1997.

[27] ———, *Cardinality, counting, and equinumerosity*, ***Notre Dame Journal of Formal Logic***, vol. 41 (2000), pp. 187–209.

[28] DAVID LEWIS, ***Parts of classes***, Blackwell, Oxford, 1991.

[29] ØYSTEIN LINNEBO, *Frege's proof of referentiality*, forthcoming in ***Notre Dame Journal of Formal Logic***.

[30] ———, *Plural quantification exposed*, ***Noûs***, vol. 37 (2003), pp. 71–92.

[31] LEOPOLD LÖWENHEIM, *Über Möglichkeiten im Relativkalkül*, ***Mathematische Annalen***, vol. 76 (1915), pp. 447–470, translated as "On possibilities in the calculus of relatives" in [38].

[32] CHARLES PARSONS, *Frege's theory of number*, 1965, reprinted in [12] and [33].

[33] ———, ***Mathematics in philosophy***, Cornell University Press, Ithaca, NY, 1983.

[34] IAN RUMFITT, *Hume's principle and the number of all objects*, ***Noûs***, vol. 35 (2001), pp. 515–41.

[35] BERTRAND RUSSELL, *Letter to Frege*, 1902, in [38].

[36] Matthias Schirn (editor), *Philosophy of mathematics today*, Oxford, Clarendon, 1998.

[37] NEIL TENNANT, *On the necessary existence of numbers*, ***Noûs***, vol. 31 (1997), pp. 307–336.

[38] Jean van Heijenoort (editor), *From Frege to Gödel*, Cambridge, MA, Harvard University Press, 1967.

[39] CRISPIN WRIGHT, ***Frege's conception of numbers as objects***, Aberdeen University Press, Aberdeen, 1983.

[40] ———, *The harmless impredicativity of $N^{=}$ (Hume's principle)*, 1998, in [36]; reprinted in [21].

[41] ———, *Response to Michael Dummett*, 1998, in [36]; reprinted in [21].

[42] EDWARD ZALTA, *Natural numbers and natural cardinals as abstract objects: A partial reconstruction of Frege's* grundgesetze *in object theory*, ***Journal of Philosophical Logic***, vol. 28 (1999), pp. 619–660.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF OSLO
POSTBOKS 1020 BLINDERN
N-0315 OSLO, NORWAY
E-mail: oystein.linnebo@filosofi.uio.no