APPENDIX: FREGE'S THEOREM AND THE ZERMELO-BOURBAKI LEMMA[40]

JOHN L. BELL

This Appendix establishes the existence of an infinite well-ordering as a (hitherto unremarked) consequence of a general version of Zermelo's Well-ordering theorem. We also indicate how this fact can be derived along 'Fregean' lines within a certain system $\mathbb{F}$ of many-sorted first-order logic whose sorts correspond to Frege's domains of objects, relations, and first and second level concepts. We show that the system of axioms we formulate within $\mathbb{F}$ constitutes a consistent fragment of Frege's original (inconsistent) system sufficient for the development of arithmetic.

We begin by specifying the basic constituents of the system $\mathbb{F}$.

Sorts (or domains)

$\mathcal{O}$ — objects

$\mathcal{B}$ — basic (first level) concepts

$\mathcal{R}$ — relations

$\mathcal{S}_b$ — second level concepts

$\mathcal{S}_r$ — second level relational concepts

Variables and Constants

Sort	*Variable*	*Constant*
$\mathcal{O}$	$x, y, z, \ldots$	$a, b, c, \ldots$
$\mathcal{B}$	$X, Y, Z, \ldots$	$A, B, C, \ldots$
$\mathcal{R}$	$\underline{X}, \underline{Y}, \underline{Z}, \ldots$	$\underline{A}, \underline{B}, \underline{C}, \ldots$
$\mathcal{S}_b$	$\utilde{X}, \utilde{Y}, \utilde{Z}, \ldots$	$\utilde{A}, \utilde{B}, \utilde{C}, \ldots$
$\mathcal{S}_r$	$\underline{\underline{X}}, \underline{\underline{Y}}, \underline{\underline{Z}}, \ldots$	$\underline{\underline{A}}, \underline{\underline{B}}, \underline{\underline{C}}, \ldots$

A *term* is a variable or a constant or one of the concept or relation or extension terms to be introduced shortly. A variable of sort $\mathcal{B}$ or $\mathcal{S}_b$ will be called a *concept variable* for brevity.

We assume the presence of an *identity* sign $=$ yielding atomic statements of the form $s = t$ where s and t are terms of the *same*

[40]Abstracted from the Appendix to William Demopoulos and John L. Bell, "Frege's theory of concepts and objects and the interpretation of second-order logic," *Philosophia mathematica* (Series III), **1** (1993) 139–56. Reprinted by kind permission of Robert Thomas, editor of that journal, and John L. Bell.

sort. On all domains except $\mathcal{O}$, $=$ is to be thought of as *intensional* equality.

We also assume the presence of a *predication* sign η yielding atomic statements of the form $s\,\eta\,t$, $(s't')\,\eta\,u$ where s is of sort $\mathcal{O}$, $\mathcal{B}$, $\mathcal{R}$ and t is of sort $\mathcal{B}$, $\mathcal{S}_b$, $\mathcal{S}_r$ respectively; and s', t' are both of sort $\mathcal{O}$ and u is of sort $\mathcal{R}$. We read '$s\,\eta\,t$' as '*s falls under t*.'

We shall assume the following comprehension scheme for concepts:

Corresponding to any formula $\Phi(x)$, $\Phi(x,y)$, $\Phi(X)$ or $\Phi(\underline{X})$ we are given a term s of sort $\mathcal{B}$, $\mathcal{R}$, $\mathcal{S}_b$, $\mathcal{S}_r$, respectively, for which we adopt as an axiom the formula

$$\forall \left\{ \begin{matrix} x \\ xy \\ X \\ \underline{X} \end{matrix} \right\} \left[\left\{ \begin{matrix} x \\ (xy) \\ X \\ \underline{X} \end{matrix} \right\} \eta\, s \leftrightarrow \Phi \left\{ \begin{matrix} x \\ xy \\ X \\ \underline{X} \end{matrix} \right\} \right].$$

We write $\widehat{x}\Phi$, $(xy)^\frown\Phi$, $\widehat{X}\Phi$, $\widehat{\underline{X}}\Phi$ for s, as the case may be. A term of the first, third and fourth types is called the *concept* (term) determined by Φ, and a term of the second type the *relation* (term) determined by Φ.

We define the relation $\equiv$ of *extensional equality* on the domains $\mathcal{B}$, $\mathcal{R}$, $\mathcal{S}_b$, $\mathcal{S}_r$ by

$$X \equiv Y \iff_{df} \forall x(x\,\eta\,X \leftrightarrow x\,\eta\,Y)$$

$$\underline{X} \equiv \underline{Y} \iff_{df} \forall x \forall y[(xy)\,\eta\,\underline{X} \leftrightarrow (xy)\,\eta\,\underline{Y}]$$

$$\underset{\sim}{X} \equiv \underset{\sim}{Y} \iff_{df} \forall X[X\,\eta\,\underset{\sim}{X} \leftrightarrow X\,\eta\,\underset{\sim}{Y}]$$

$$\underline{\underline{X}} \equiv \underline{\underline{Y}} \iff_{df} \forall \underline{X}[\underline{X}\,\eta\,\underline{\underline{X}} \leftrightarrow \underline{X}\,\eta\,\underline{\underline{Y}}].$$

Clearly concepts are determined uniquely by formulas up to extensional equality. We assume that $\mathbb{F}$ contains

- a term e such that $e(\mathfrak{X})$ is well-formed and of sort $\mathcal{O}$ for any concept variable $\mathfrak{X}$;
- a predicate symbol E such that $E(\mathfrak{X})$ is well-formed for any concept variable $\mathfrak{X}$.

We finally assume the axioms

(1) $\forall\mathfrak{X}\forall\mathfrak{Y}[E(\mathfrak{X}) \wedge E(\mathfrak{Y}) \rightarrow [e(\mathfrak{X}) = e(\mathfrak{Y}) \leftrightarrow \mathfrak{X} \equiv \mathfrak{Y}]]$
(2) $\forall\mathfrak{X}\forall\mathfrak{Y}[E(\mathfrak{X}) \wedge \mathfrak{X} \equiv \mathfrak{Y} \rightarrow E(\mathfrak{Y})]$

where in both (1) *and* (2) $\mathfrak{X}$ *and* $\mathfrak{Y}$ *are concept variables of the same sort.*

If we think of $e(\mathfrak{X})$ as an object *representing* $\mathfrak{X}$, Axiom 1 above expresses the idea that *extensional equality* of any concepts satisfying E is equivalent to *identity* of their representing objects. That is, for any concept $\mathfrak{X}$ satisfying E, $e(\mathfrak{X})$ may be regarded as the *extension* of $\mathfrak{X}$. And the predicate E itself represents the property of *possessing an extension*. For these reasons Axiom 1 will be called the *Axiom of Extensions*. As for Axiom 2, it states the reasonable requirement that any concept extensionally equivalent to a concept possessing an extension *itself* possesses one (that is, $\equiv$ is a *congruence* relation with respect to E).

A straightforward Russell type argument in $\mathbb{F}$ enables us to infer $\neg\forall\mathfrak{X}E(\mathfrak{X})$,[41] that is, *not every concept possesses an extension*. This being the case, what concepts do we need to (consistently) assume possess extensions in order to enable an infinite well-ordering to be constructed? It was Frege's remarkable *discovery* that for this it suffices to assume just that extensions be possessed by the members of a certain class of simple and natural *second-order* concepts—those that, following Boolos,[42] we shall term *numerical*.

Numerical concepts are defined as follows. First, we formulate the relation $\approx$ of equinumerosity or equipollence on $\mathcal{B}$ as usual:

$$\begin{aligned} X \approx Y \Longleftrightarrow_{df} \exists \underline{Z}[&\forall x \forall y[(xy)\,\eta\,\underline{Z} \to x\,\eta\,X \wedge y\,\eta\,Y] \\ &\wedge \forall x \forall y \forall z[(xy)\,\eta\,\underline{Z} \wedge (xz)\,\eta\,\underline{Z} \to y = z] \\ &\wedge \forall x[x\,\eta\,X \to \exists y\,(xy)\,\eta\,\underline{Z}] \\ &\wedge \forall y[y\,\eta\,Y \to \exists x\,(xy)\,\eta\,\underline{Z}] \end{aligned}$$

With any basic concept X we associate the second level concept

$$\|X\| =_{df} \widehat{Y}\,[X \approx Y].$$

Concepts of the form $\|X\|$ are called *numerical*.

[41] To be explicit, define $A =_{df} \hat{x}[\forall X[e(X) = x \wedge E(X) \to \neg x\,\eta\,X]]$. Then $\neg E(A)$ is inferrable in $\mathbb{F}$.

[42] "The standard of equality of numbers," Chapter 8, below.

If we assume that every numerical concept possesses an extension (i.e., $\forall X E(\|X\|)$, then the extension

$$|X| =_{df} e(\|X\|)$$

is called the (cardinal) *number* of X. Objects of the form $|X|$ are called (cardinal) *numbers*. Under these assumptions it is easy to derive *Hume's principle*, viz.

$$\forall X \forall Y\, [X \approx Y \leftrightarrow |X| = |Y|\,].$$

We shall call a concept X (Dedekind) *infinite* if $\exists Y\, [Y \subsetneqq X \wedge X \approx Y]$, where $Y \subsetneqq X$ of course stands for $\forall x (x\, \eta\, Y \rightarrow x\, \eta\, X) \wedge Y \not\equiv X$. Objects of the form $|X|$ with X infinite are called *infinite numbers*.

We are going to show how, in $\mathbb{F}$, the existence of an infinite well-ordering (i.e. an infinite well-ordered concept) may be derived as a special case of a general set-theoretic result—formulable and provable in $\mathbb{F}$—which is normally used to derive Zermelo's Well-ordering theorem. In its original form this result is what we shall call the

Zermelo-Bourbaki lemma. [43] *Let E be a set, $\mathcal{F}$ a family of subsets of E and $p : \mathcal{F} \rightarrow E$ a map such that $p(X) \notin X$ for all $X \in \mathcal{F}$. Then there is a subset M of E and a well-ordering $\leqslant$ of M such that, writing S_x for $\{y : y < x\}$,*

(i) $\forall x \in M\, [S_x \in \mathcal{F} \wedge p(S_x) = x]$
(ii) $M \notin \mathcal{F}$.

Bourbaki employs this result to construct an elegant derivation of Zermelo's Well-ordering theorem from the Axiom of Choice. In the present context, however, it will be used to produce an equally elegant proof of what we shall call, following a suggestion of Boolos,

[43] Lemma 3, § 2, Ch. 3 of N. Bourbaki, *Théorie des ensembles*, 2nd ed. Paris: Hermann, 1963. Bourbaki's proof is a generalization of Zermelo's argument for his Well-ordering theorem in his "Proof that every set can be well-ordered" (1904) in Jean van Heijenoort, ed., *From Frege to Gödel: a sourcebook in mathematical logic, 1879–1931*, Cambridge: Harvard University Press, 1967, Stefan Bauer-Mengelberg, tr.

Frege's theorem. *Suppose given a set E and a map $n: PE \to E$ such that*

$$(*) \qquad \forall X \subseteq E \, \forall Y \subseteq E \, [n(X) = n(Y) \leftrightarrow X \approx Y].$$

Then E has an infinite well-ordered subset.

Proof. We apply the Zermelo-Bourbaki lemma with $\mathcal{F}$ the family of all subsets X of E for which $n(X) \notin X$ and p the map n. We obtain $M \subseteq E$ and a well-ordering $\leqslant$ of M such that *(i)* $n(S_x) = x$ for all $x \in M$, *(ii)* $n(M) \in M$. Writing m for $n(M)$ we have $m \in M$ by *(ii)*, whence $n(S_m) = m = n(M)$ by *(i)*. Condition $(*)$ now yields $S_m \approx M$. Since $m \notin S_m$, S_m is a proper subset of M and it follows that the latter is infinite. □

Now both of these results can be translated into and proved within $\mathbb{F}$. Carrying this out for the Zermelo-Bourbaki lemma yields the

Zermelo-Bourbaki lemma in $\mathbb{F}$. *Let $\underset{\sim}{S}$ be any second level concept with respect to which $\equiv$ is a congruence relation and t a term such that $t(X)$ is an object for all basic concepts X and satisfies*

$$\forall X \forall Y \left[X \equiv Y \wedge X \, \eta \, \underset{\sim}{S} \to t(X) = t(Y)\right]$$

$$\forall X \left[X \, \eta \, \underset{\sim}{S} \to \neg t(X) \, \eta \, X\right].$$

Then there is a relation R such that R is a well-ordering and, writing M for its field, and R_x for $\widehat{y}\,[(yx) \, \eta \, R \wedge y \neq x]$,

(i) $\forall x \left[x \, \eta \, M \to R_x \, \eta \, \underset{\sim}{S} \wedge t(R_x) = x\right]$

(ii) $\neg M \, \eta \, \underset{\sim}{S}$.

In the case of Frege's theorem, the same process yields

Frege's theorem in $\mathbb{F}$. *Suppose that every numerical concept has an extension. Then there exists an infinite well-ordered concept and hence an infinite number.*

Since, as is well known, Frege's original system in the *Grundgesetze* was inconsistent, we should assure ourselves that the axioms of

$\mathbb{F}$, together with the hypothesis of Frege's theorem—that every numerical concept has an extension—are consistent. The easiest way to see this is by noting that the following set-theoretic interpretations yield a model of the axioms of $\mathbb{F}$ in which the hypothesis of Frege's theorem holds. To wit, interpret $\mathcal{O}$ as $\omega+1$, $\mathcal{B}$ as $P(\omega+1)$, $\mathcal{R}$ as $P((\omega+1)\times(\omega+1))$, $\mathcal{S}_b$ as $PP(\omega+1)$, $\mathcal{S}_r$ as $PP((\omega+1)\times(\omega+1))$, E as the subset of $PP(\omega+1)$ consisting of all elements of the form $\|X\| =_{df} \{Y \in P(\omega+1) : X \approx Y\}$ with $X \in P(\omega+1)$, and e as the map $PP(\omega+1) \to \omega+1$ which sends each $\|X\|$ with $X \in P(\omega+1)$ to its cardinality $|X|$ ($\in \omega+1$) and everything else to 0. Thus the axioms of $\mathbb{F}$ — together with the hypothesis of Frege's theorem — may be regarded as a consistent fragment of Frege's original system.