Math. Log. Quart. 40 (1994) 27 – 30

Mathematical Logic
Quarterly

Fregean Extensions of First-Order Theories

John L. Bell

Department of Philosophy, University of Western Ontario,
London, Ontario, Canada N6A 3K7

Abstract. It is shown by Parsons [2] that the first-order fragment of Frege's (inconsistent) logical system in the *Grundgesetze der Arithmetic* is consistent. In this note we formulate and prove a stronger version of this result for arbitrary first-order theories. We also show that a natural attempt to further strengthen our result runs afoul of Tarski's theorem on the undefinability of truth.

Mathematics Subject Classification: 03B10.

Keywords: Fregean theory, Frege's abstraction scheme, Fregean extension, Conservative extension.

§1. Fregean theories

We shall call a theory T in a (countable) first-order language $\mathcal{L}$ *Fregean* if corresponding to each formula $A(x)$ with one free variable x there is a constant symbol c_A of $\mathcal{L}$ such that, for any such formulas $A(x)$, $B(x)$, the sentence

(1) $\quad \forall x[A(x) \leftrightarrow B(x)] \leftrightarrow c_A = c_B$

is provable in T.

When T is the first-order fragment of FREGE's system as presented in [2], the scheme of sentences (1) is the first-order version of FREGE's abstraction principle for predicates (concepts) with the constant c_A playing the role of the "extension" ("Umfang") of the predicate A. We shall call this scheme *Frege's abstraction scheme*.

Our main result is the

Theorem. *If T has no finite models, then it has a conservative Fregean extension.*

Proof. We begin by introducing new constants which we shall call *special constants*: each of these will be assigned a natural number called its *level*. The special constants are defined by induction on levels as follows. Suppose that special constants of all levels $< n$ have been defined. Let $A(x)$ be a formula using symbols of $\mathcal{L}$ and just the constants of level $< n$. If $n > 0$, suppose also that $A(x)$ contains at least one special constant of level $n - 1$. Then the symbol c_A is a special constant of level n called *the special constant assigned to $A(x)$*.

Let $\mathcal{L}^*$ be the language obtained by adding to $\mathcal{L}$ all special constants of all levels. Each formula $A(x)$ of $\mathcal{L}^*$ is then assigned a unique special constant: its level is the least number exceeding the levels of all special constants occurring in $A(x)$. If $A(x)$, $B(x)$ are formulas of $\mathcal{L}^*$, the sentence (1) is called the *Fregean axiom for* (c_A, c_B).

Let T^* be the theory in $\mathcal{L}^*$ obtained by adding all the Fregean axioms to T. Clearly T^* is Fregean (in $\mathcal{L}^*$).

Suppose now that T has no finite models. We claim that T^* *is a conservative extension of* T. To prove this it suffices to show that any model M of T can be expanded to a model M^* of T^* with the same domain.

To obtain M^* we provide interpretations in (the domain of) M of all special constants recursively as follows. Suppose that interpretations in M have been assigned to all special constants of level $< n$ in such a way that

1) all the corresponding Fregean axioms are satisfied;

2) the set M' of elements of M which are *not* interpretations of special constants of level $< n$ has cardinality $|M|$ (necessarily infinite).

Since M' is infinite it can be partitioned into two subsets M'' and M''' each of cardinality $|M'| = |M|$. Moreover, since M is infinite, the number of subsets of the model definable using formulas of $\mathcal{L}$ involving only special constants of level $< n$ is $\leq |M| = |M''|$. Let f be an injection of this collection of sets into M''. We define an interpretation of each special constant c_A of level n (in such a way as to preserve the truth of 1) and 2)) as follows:

a) If for some special constant c_B of level $< n$, $A(x)$ and $B(x)$ define the same subset of M, then c_A is assigned the same interpretation as that which has (by induction hypothesis) already been assigned to c_B.

b) Otherwise, c_A is assigned the interpretation $f(X)$, where X is the subset of M defined by $A(x)$.

(Note that in case a) there may be more than one formula of level $< n$ defining the same subset of M as $A(x)$: but condition 1) of the induction hypothesis implies that all the corresponding special constants receive the *same* interpretation. So the stipulation in case a) is consistent.)

This recipe furnishes interpretations in M of all the special constants. It is now easy to check that all the Fregean axioms are true in the resulting model M^*, which is therefore an expansion of M to a model of T^*.

This completes the proof of the theorem. □

Corollary. *If T has an infinite model, then it has a consistent Fregean extension.*

Proof. Let T' be obtained from T by adding as axioms the sentences $d_m = d_n$ for $m \neq n$, where $\{d_m : m \in \omega\}$ is a set of new distinct constants. Since T has an infinite model, T' is consistent. Clearly T' has no finite models, so the theorem implies that T' has a conservative Fregean extension; this latter is a consistent Fregean extension of T. □

We note that the assertions of the Theorem and its Corollary can be reversed: that is, any theory with a conservative (respectively consistent) Fregean extension has no finite models (respectively has an infinite model). This follows immediately from the observation that

Any model of a Fregean theory is infinite.

To prove this, take any Fregean theory T and define the sequence of formulas $A_0(x), A_1(x), \dots$ recursively as follows. First, set $A_0(x) \equiv (x = x)$ and for $n \geq 1$,

$$A_n(x) \equiv \bigwedge_{i<n} x \neq c_{A_i}.$$

For simplicity write c_n for c_{A_n}. To prove the result it suffices to show that $\vdash_T c_m \neq c_n$ for $m < n$. We do this by induction on n. Suppose that $n \geq 1$ and $\vdash_T c_i \neq c_m$ for all $i < m < n$. Then if $m < n$, it follows that

$$\vdash_T A_m(c_m) \wedge \neg A_n(c_m)$$

so that

$$\vdash_T \neg \forall x[A_m(x) \leftrightarrow A_n(x)].$$

The Fregean axiom for (c_m, c_n) now yields $\vdash_T c_m \neq c_n$, completing the induction step and the proof.

§2. Strengthening the abstraction scheme

It is well known that the inconsistency in the *Grundgesetze der Arithmetik* arises from the fact that FREGE assumes the full second-order version of the abstraction scheme. For our purposes this may be formulated as:

$$(2) \qquad \forall P \forall Q \left[\forall x[P(x) \leftrightarrow Q(x)] \leftrightarrow c(P) = c(Q)\right],$$

where P, Q are predicate variables and c is a function assigning individual (lowest level) terms to predicate variables. To see that (2) is inconsistent we suppose that $\mathcal{S}$ is a second-order theory in which (2) is provable, and define the formula $A(x)$ by

$$A(x) \equiv \exists P[x = c(P) \wedge \neg P(x)].$$

Then

$$\begin{aligned} \vdash_{\mathcal{S}} A(c(A)) &\leftrightarrow \exists P[c(A) = c(P) \wedge \neg P(c(A))] \\ &\leftrightarrow \exists P\left[\forall x[A(x) \leftrightarrow P(x)] \wedge \neg P(c(A))\right] \\ &\leftrightarrow \neg A(c(A)). \end{aligned}$$

Thus the attempt to strengthen FREGE's abstraction scheme by allowing second-order quantification *and* functional dependence of "extensions" as predicates leads to inconsistency, because it runs afoul of RUSSELL's paradox, or, equally, CANTOR's diagonalization argument. It is of interest to note that inconsistency can still arise even *without* second-order quantification, provided we insist that the functional dependence of "extensions" on predicates be *definable*. In this case, as we shall see, the inconsistency arises for what seems to be quite a different reason, namely, as a consequence of TARSKI's theorem on the undefinability of truth. (A related point is made in [1].)

Let us call a theory T *Tarskian* if enough of the syntax of its language $\mathcal{L}$ can be encoded within T to ensure that T is subject to TARSKI's undefinability theorem. This means that, if T is consistent, then, writing $\ulcorner A \urcorner$ for the code in $\mathcal{L}$ of a formula A, there is no formula $W(x)$ of $\mathcal{L}$ such that

$$\vdash S \leftrightarrow W(\ulcorner S \urcorner)$$

for every sentence S of $\mathcal{L}$. Note that both first-order arithmetic and set theory are Tarskian.

Now, let $\tau(x)$ be a term of $\mathcal{L}$ and consider the scheme

$$(3) \qquad \forall x[A(x) \leftrightarrow B(x)] \leftrightarrow \tau(\ulcorner A \urcorner) = \tau(\ulcorner B \urcorner),$$

obtained from FREGE's abstraction scheme by replacing c_A and c_B by their "definable" counterparts $\tau(\ulcorner A \urcorner)$ and $\tau(\ulcorner B \urcorner)$. We conclude with the

Theorem. *If T is Tarskian, and there is a term τ such that* (3) *is provable in T for all formulas $A(x)$, $B(x)$, then T is inconsistent.*

Proof. Assume that T satisfies the specified conditions. Given a sentence S, define

$$A(x) \equiv S \wedge (x = x) \quad \text{and} \quad B(x) \equiv (x = x).$$

Then

$$\vdash_T \forall x[A(x) \leftrightarrow B(x)] \leftrightarrow S,$$

so that, by the provability of (3) in T,

$$\vdash_T S \leftrightarrow \tau(\ulcorner A \urcorner) = \tau(\ulcorner B \urcorner).$$

But the representability of the syntax of $\mathcal{L}$ within T implies that there is a formula $W(x)$ of the language of T such that $W(\ulcorner S \urcorner)$ is $\tau(\ulcorner A \urcorner) = \tau(\ulcorner B \urcorner)$ (for arbitrary S). Then $\vdash_T S \leftrightarrow W(\ulcorner S \urcorner)$ for any sentence S; since T is Tarskian, its inconsistency follows. □

References

[1] ACZEL, P., Frege structures and the notions of propositions, truth and set. In: The Kleene Symposium (J. BARWISE, H. J. KEISLER, and K. KUNEN, editors), North-Holland Publ. Co., Amsterdam, 1980, pp. 31 – 59.

[2] PARSONS, T., On the consistency of the first-order portion of Frege's logical system. Notre Dame J. Formal Logic **28** (1987), 161 – 168.

(Received: May 22, 1993)