SOME OBSERVATIONS RELATED TO FREGE'S WAY OUT *

MICHAEL D. RESNIK

In this note I shall make some observations concerning both the original and repaired systems presented by Frege in his *Grundgesetze der Arithmetik*. These in turn lead to general considerations concerning related axiom systems and contemporary comparative set theory. I hope that my remarks will be useful to others — as they were to me — for obtaining some insight into Frege's and current systems.

1. *A generalization of Frege's derivation of the Russell contradiction.*

Out of historical interest let us derive the Russell paradox — much like Frege did — in a second order functional calculus. (1) Following him we take as our class axiom

$$\text{V. } \hat{x}F(x) = \hat{y}G(y) .\equiv. (x)\ [F(x) \equiv G(x)];$$

but instead of considering the concept *class not falling under its own concept,* we make the following generalization: Let ψ be a one-one function, that is, assume

$$\text{A. } (x)\ (y)\ [\psi(x) = \psi(y) \supset x = y].$$

Then, of course, ψ has an inverse. Let us call the values of ψ for classes as arguments ψ-classes. Then consider the concept ψ*-class not falling under the concept of its inverse* — abbreviated 'R' and symbolized by '(EP) $[x = \psi(\hat{y}P(y)).\ \sim P(x)]$'. The ψ-class of its extension, $\psi(Z)$, is a generalization of the Russell class, so it should not surprise us that we can use it to derive a contradiction.

1.1 $\psi(Z)$ is a ψ-class not falling under the concept of its inverse. For if it is not, then it does not fall under the concept of Z. But then it is a ψ-class not falling under the concept of its inverse.
Formally:

* I am grateful to John Vickers for his perceptive comments on earlier drafts of this paper.

(1) G. Frege, *Grundgesetze der Arithmetik II*, p. 256-257.

1. $(P)[\psi(Z)=\psi(\hat{y}P(y)) \supset P(\psi(Z))] \supset.\ \psi(Z)=\psi(Z) \supset.\ R(\psi(Z))$.
2. $\sim R(\psi(Z)) \supset R(\psi(Z))$. (1) by def. and identity theory
3. $R(\psi(Z))$. (2)

[The antecedent clause of (1) is a quantificational transform of '$\sim R(\psi(Z))$'; the consequent is the result of instantiating 'P' by 'R' and abbreviating.]

1.2 $\psi(Z)$ is not a ψ-class not falling under the concept of its inverse. For by 1.1, it is a ψ-class not falling under the concept of its inverse; thus it falls under the concept of Z, i.e., the concept of its inverse. Thus $\psi(Z)$ is not a ψ-class not falling under the concept of its inverse. Formally:

1. $\psi(Z)=\psi(\hat{y}G(y)) \supset Z = \hat{y}G(y)$ (A)
2. $Z = \hat{y}G(y) \supset (x)\ [R(x)\equiv G(x)]$ (V)
3. $\psi(Z)=\psi(\hat{y}G(y)) \supset.\ R(\psi(Z)) \supset G(\psi(Z))$ (1), (2)
4. $\psi(Z)=\psi(\hat{y}G(y)) \supset G(\psi(Z))$ (3), 1. 1. (3)
5. $\sim R(\psi(Z))$ (4) by generalization, quantificational transformation, and abbreviation.

Obviously Frege's system is inconsistent since we can take ψ as the identity function and we get Russell's contradiction. Also note that we only need to use our class axiom in 1.2. (2)

The derivations just given can also be easily modified (by replacing 'ψ' by 'f' to prove in Frege's system:

1.1.a. $(f)R_f(f(Z_f))$.
1.2.a. $(f)[(x)(y)[f(x) = f(y) \supset x = y] \supset \sim R_f(f(Z_f))]$. (3)

2. *First order systems.*

Although this contradiction has been derived in a second order system in which both classes and functions are assumed explicitly, it holds for the first order correlate of Frege's system too. Let us assume first order quantification theory with identity and take, as the class axiom, the axiom of comprehension

B. $\ulcorner(E\alpha)(\beta)[\beta \varepsilon \alpha \equiv \Phi]$,$\urcorner$ where Φ does not contain α free.

(2) The derivations just given are minor modifications of Frege's.
(3) '$R_f(x)$' is '$(EP)\ [x = f(yG(y)).\ \sim F(x)]$', where 'f' is a free function variable.

It is plausible to reguard instances of (B) such as

$$(Ey)(x)[x \varepsilon y \equiv \sim (x \varepsilon a)], \quad (Ey)(x)[x \varepsilon y \equiv x = a]$$

as introducing functions deined on classes — in this case the complement of a, $\bar{a}$, and the unit class of a, ι a, respectively. We shall say that any instance of (B), where Φ contains only β and some other variable (different from α, of course) free, introduces a class function in the restricted sense. These functions need not be functions in a more general sense. E.g., as we shall see Quine develops functions in the general sense as classes or ordered pairs; class functions introduced by his axiom of comprehension do not all have classes of ordered pairs answering to them: the values of every function in the general sense in Quine's system are elements; x is an element if and only if x is.

In the present system any functions in the restricted sense which can be proved to be one-one may be used to define a paradoxical class.

E.G.

1.	$\iota x = \iota y \supset x = y.$	(B)
2.	$(Ey)(x)[x \varepsilon y \equiv (Ez)[x = \iota z . \sim (x \varepsilon z)]]$	(B)
3.	$(Ey)[\iota y \varepsilon y \equiv (Ez)[\iota y = \iota z . \sim (x \varepsilon z)]]$	(2)
4.	$(Ey)[\iota y \varepsilon y \equiv \sim (\iota y \varepsilon y)]$	(1)(3)

3. *Frege's way out.*

Frege avoided the Russell contradiction by replacing (V) by

V'. $\hat{x}F(x) = \hat{y}G(y) . \equiv . \ (z)[z \neq \hat{x}F(x) \supset . z \neq \hat{y}G(y) \supset . F(z) \equiv G(z)]$. (4)

Thus his repaired system is trivially consistent, since (V') has a model in any domain with exactly one individual. Moreover, step (2) of 1.2 is now blocked.

Nonetheless, it is known that if very simple axioms such as '$(Ex)(Ey)[x \neq y]$' are added to Frege's new system, new contradictions may be derived. (5)

(4) FREGE, *op.cit.*, p. 262-263.

(5) See: SOBOCINSKI, «L'analyse de l'antinomie Russelliene par Lesniewski», *Methodes*, 1-2, 1949-1950. QUINE, «On Frege's way out», *Mind*, *64*, 1955. GEACH, «On Frege's way out», *Mind*, *65*, 1956.

In the repaired system (B) no longer holds, and is replaced (in the first order correlate) by

B'. $\ulcorner(E\alpha)(\beta)[\alpha \neq \beta \supset . \beta \varepsilon \alpha \equiv \varphi]$, [1] where is as before. Then we can prove '$x \neq \iota x . \iota x = \iota y . \supset x = y$'; however, without first deriving a contradiction first (using additional assumptions) and using ($p . \sim p . \supset q$', I have not been able to drop the condition '$x \neq \iota x$.' This seems to be true generally of functions in the restricted sense which were one-one in the system containing (B). Nonetheless, according to Sobocinski and Geach, Lesniewski succeeded in using one of these former one-one functions — ιx — to derive a contradiction from (B') and '$(Ex)(Ey)(Ez)[x \neq y . x \neq z . y \neq z]$'. (6)

4. *A new contradiction.*

By making a stronger assumption than those previously made, namely, by assuming the existence of one-one functions without fixed points, we can easily modify the derivations of 1.1 and 1.2 to obtain a new inconsistency in the repaired system. For '$(f)R_f(f(Z_f))$' can be proved just as before, and by the obvious changes in 1.2. we obtain

$$(f)[(x)(y)[f(x) = f(y) \supset x = y] \supset . (x)[f(x) \neq x] \supset \sim R_f(f(Z_f))].$$

Then if we take

C. $(Ef)[(x)(y)[f(x) = f(y) \supset x = y] . (x)[f(x) \neq x]]$

as an axiom the contradiction in the system containing it and (V') is immediate.

One could rightly object that (C) is an overly strong assumption to make. Yet even if we give it up, we have shown Frege's repaired system to lack standard models in any domain with at least two individuals. For (C) is true in such models. This result is far weaker

(6) J. M. Bartlett has remarked that any one-one function gives rise to a paradoxical class in Frege's system and that only the most simple of these are avoided in the repaired system. Since he does not supply the details, I do not know whether he had anticipated the observations just made or whether he also knew those to be made in the next section. My discoveries were made independently of his — until I had made them I failed to comprehend the remark in question. See J. M. Bartlett, *Funktion und Gegenstand*, p. 65-67 (Dissertation, Munich, 1961).

than that already obtained by Quine [7] — the restriction to standard models may be dropped — but the technique used to obtain it admits of interesting generalizations.

5. *Generalizations and limitations.*

Suppose that we again consider class axioms formulated in a second order functional calculus, but instead of Frege's axiom let us take one of the form $\hat{x}F(x)=\hat{y}G(y).\equiv.(x)[x\neq a\supset.F(x)\equiv G(x)]$, where 'a' is some constant, e.g., 'Λ.' Clearly this system has no standard models in infinite domains. For however 'a' may be interpreted, one- one functions may be found on such domains which do not map anything into the object to which 'a' refers under its interpretation. Is is easily verified that in this system we can also prove

$$(f)R_f(f(Z_f))$$
$$(f)[(x)(y)[f(x)=f(y)\supset x=y]\supset.(x)[f(x)\neq a]\supset\sim R_f(f(Z_f))].$$

Thus the system has no models under any interpretation which makes true:

$$(Ef)[(x)(y)[f(x)=f(y)\supset x=y].(x)[f(x)\neq a]].$$

Although the same argument also works if '$x\neq a$' is replaced by '$x\neq a_1. \ldots.x\neq a_n$', the technique involved no longer applies when '$x\neq a$' is replaced by an arbitrary condition 'P(x).' We no longer possess enough information to conclude anything about its various interpretations.

6. A '*fallacious*' *argument.*

We do know, however, that if both

D. $\hat{x}F(x)=\hat{y}G(y).\equiv.(x)[P(x)\supset.F(x)\equiv G(x)]$ and
E. $(Ef)[(x)(y)[f(x)=f(y)\supset x=y].(x)P(f(x))]$, where 'P' is a class

theoretic predicate, are provable in a given system then it is inconsistent. Thus it might be profitable to see if any basic or elemen-

(7) QUINE, *op.cit.*

tary 'facts' can be used to prove to (E). If some of these could not be used together with (D), we might want to discount formulas of its form.

Neither '(x)P(x)' nor '(x) ~ P(x)' should be theorems of a system containing (D) as an axiom; the first clearly reinstates the Russell contradiction; the second renders part of (D) useless. Thus we should wish to have both '(Ex)P(x), and '(Ex) ~ P(x)' as theorems in an adequate system and to be able to formalize the following argument: Let X be the class of non-P. It is neither empty nor universal. Thus there is a superclass Y of X such that $Y \neq X$, whose cardinality is the first aleph greater than that of X. But then there is a one-one function Φ defined on Y,mapping Y into Y, which takes no member of Y into a member of X. Define Φ' by:

if x is in Y, then $\Phi'(x) = \Phi(x)$
if x is not in Y, then $\Phi'(x) = x$

Φ' is one-one and $(x)P(\Phi'(x))$. Thus (E) is proved and our system, since it contains (D), is inconsistent.

In spite of the intuitive soundness of this argument, it would certainly be fallacious to offer it as a consideration against (D). For even if we translate it into first order terms, it does not appear that we shall succeed in formalizing it in any of the currently competing set theories. E.g., we have permitted ourselves to build arbitrarily large classes; we have not distinguished between classes and sets or elements; we have defined functions on the whole universe of discourse. In each case we have violated the principles of some set theory; thus it is likely that our argument will break down once we try to formalize it in a set theory not known to be inconsistent.

Indeed — to take a concrete case — we can be certain of this with respect to Quine's *Mathematical Logic* (relative to our confidence in that system), because otherwise it would be demonstrably inconsistent. For if we replace 'P' by '$x \varepsilon V$' and translate (D) into first order terms, we obtain

F. $[(E\alpha)(\beta)(\beta \varepsilon V \supset . \beta \varepsilon \alpha \equiv \Phi)]$, where Φ is as above, which is implied by Quine's *202; (8) thus if we would prove in Quine's system that there is a one-one function ψ (in either the general or restricted sense) such that

$(x)[\psi(x) \varepsilon V]$,

the system would be inconsistent.

(8) Quine, *Mathematical Logic*. Revised edition, 1958, p. 162.

Nonetheless, the obvious ways to formalize the proof given above have to break down. For according to Quine's development of functions in the general sense, each function has the null class a value for any non-element as an argument; even the identity function developed thus is not one-one. (9) It is interesting then that '(x) [f'x ε V]' is a theorem. (10) On the other hand, attempts to use functions in the restricted sense appear to be blocked too. For symmetry let us consider the complement function, $\bar{x}$. We have the theorem '(x)(y)[$\bar{x}=\bar{y} \supset x=y$]'; but since the complement of a non-element is a non-element, our other demand will not be met. (11) Attempts to use other functions such as the unit class function appear to be obstructed also.

7. *Conclusion.*

The moral of our story must be a very well known one: set theory is now comparative; we can no longer assume that 1) even elementary arguments from Cantor's theory will go through in current systems intact, or, 2) the arguments acceptable in one system will be acceptable in another. Indeed, our 'fallacious' argument will be seen as fallacious for reasons which in turn depend on one's set theory — Quine would attack it for failing to distinguish classes from elements, Zermelo would not object to this but would protest against our failing to limit the size of our classes. Yet possibly both would admit the merits of the other's position. When faced with a situation like this one can sympathize with, if not accept, Lesniewski's remark: «An unintuitive mathematics contains no effective remedy for the troubles of intuition». (12)

Michael D. Resnik

Air Force Cambridge Research Laboratories (OAR)
Bedford, Massachusetts
and
Department of Philosophy,
University of Hawaii,
Honolulu, Hawaii.

(9) *Ibid.*, p. 223, p. 229.
(10) *Ibid.*, p. 234.
(11) *Ibid.*, p. 181-183.
(12) Lesniewski, «Grundzüge eines neuen Systems der Grundlagen der Mathematik», *Fundamenta Mathematicae* XIX, 1929, p. 7.