DIAGONALIZATION, THE LIAR PARADOX, AND THE INCONSISTENCY OF THE FORMAL SYSTEM PRESENTED IN THE APPENDIX TO FREGE'S *GRUNDGESETZE*: *VOLUME II*

ROY T COOK
University of Minnesota & University of St Andrews

Abstract
The *Liar Paradox* is constructed within Frege's *Grundgesetze* using a variant of Gödel's diagonalization lemma. The particular instance of Basic Law V that triggers the *Liar paradox* is identified, and it is observed that this is exactly the principle that Frege himself identified as the root of *Russell's paradox* in the appendix to Volume II of the *Grundgesetze*. Unfortunately, the amended version of Basic Law V which Frege suggests as a patch to his system blocks neither the derivation of the diagonalization theorem nor the construction of the Liar paradox.

1 DIAGONALIZATION IN THE GRUNDGESETZE

The standard description of the formal system presented in Gottlob Frege's *Grundgesetze* is typically described as follows: Frege's system amounts to nothing more than (or, more carefully, is equivalent to) higher-order logic plus the inconsistent *Basic Law V*:

BLV: $(\forall X)(\forall Y)[(\S(X) = \S(Y)) = (\forall z)(Xz = Yz)]$[1]

Such a description is incorrect, however: There are a number of aspects of Frege's logic that differentiate it from standard higher-order systems such

[1] Here, and below, I use modern symbolism instead of Frege's two-dimensional notation, primarily for typographical convenience. All proofs, etc., can be straightforwardly translated into Frege's original formalism. Particular attention should be paid to the use of identity, since in Frege's system identity holding between two statements (i.e. names of truth values) is roughly equivalent to our biconditional.

as those studied in Shapiro [1991]. An exploration of these differences, and how they affect the proof-theoretic strength of Frege's system and various natural modifications of it, promises to shed light both on Frege's logical thought, and on the roots of a number of central paradoxes

The first difference between the system of Frege's *Grundgesetze* and modern formal systems is that Frege treats statements (or, more carefully, what *we* would think of as statements) as names of truth values. Thus, the connectives represent, quite literally, truth-functions, and quantification into sentential position is allowed. (These are first-order quantifiers, distinguishing Frege's approach from higher-order logics which allow for second-order quantification into sentential position, interpreting such quantifiers as ranging over 'concepts' of zero arity). For example, the *Grundgesetze* analogue of:

$$(\exists x)(\sim x)$$

is both well-formed and a theorem in Frege's *Grundgesetze*.

Once we realize that the quantifiers of the *Grundgesetze* range over not just value ranges and other mathematical (and perhaps non-mathematical) objects, but also over truth values, the second aspect of Frege's system which will be of interest becomes apparent. Frege's language contains a falsity predicate:

$$x = \sim(\forall y)(y = y)$$

In other words, an object is the false if and only if it is identical with the truth value denoted by:

$$\sim(\forall y)(y = y)$$

Note, however, that the falsity predicate (and the corresponding truth predicate) constructed within the *Grundgesetze* operates a bit differently from the manner in which falsity predicates (and corresponding truth predicates) operate within formal systems today: The intended extension of a truth predicate is normally understood today to be the class of all true statements (or, if working within arithmetic and using Gödel coding, the class of all natural numbers that code true statements). In Frege's system, however, the falsity predicate is the predicate that holds of the false – that is, of the unique object that is the referent of all false statements.

Thus, within the *Grundgesetze*, we can quantify over statements and we can construct a falsity predicate. The natural next question to ask is whether the *Liar Paradox* can be constructed within Frege's system. The answer is "Yes". To do so, we first define a diagonalization relation as follows:

$$\text{Diag}(x, y) = (\exists Z)(y = \S Z \wedge x = Z(y))$$

"Diag" holds between x and y if and only if y is the value-range of some concept Z and x is the truth value obtained by applying Z to y – that is, by applying Z to the value-range of Z. We can now prove the following version of diagonalization:

> **Theorem 1**: In the *Grundgesetze*, for any predicate $\Phi(x)$, there is a sentence G such that:
>
> $$\Phi(G) = G$$
>
> is a theorem.

Proof: Given $\Phi(x)$, let:

$$F(y) = (\exists x)(\text{Diag}(x, y) \wedge \Phi(x))$$
$$G = F(\S F)$$

The following are provably equivalent in the *Grundgesetze*:

(1) $\Phi(G)$
(2) $\Phi(F(\S F))$
(3) $(\forall x)(F(x) = F(x)) \wedge F(\S F) = F(\S F) \wedge \Phi(F(\S F))$
(4) $(\exists Z)((\forall x)(F(x) = Z(x)) \wedge Z(\S F) = Z(\S F) \wedge \Phi(Z(\S F)))$
(5) $(\exists Z)(\S F = \S Z \wedge Z(\S F) = Z(\S F) \wedge \Phi(Z(\S F))$
(6) $(\exists x)(\exists Z)(\S F = \S Z \wedge x = Z(\S F) \wedge \Phi(x))$
(7) $F(\S F)$
(8) G

[(1) and (2) are equivalent by the definition of G, (2) and (3) by logic, (3) and (4) by logic, (4) and (5) by BLV, (5) and (6) by logic, (6) and (7) by the definition of F, and (7) and (8) by the definition of G.]

The basic idea underlying this proof is that we can 'fake' the standard proof of diagonalization (see e.g., Boolos, Burgess & Jeffrey [2007], Chapter 17, pp.220–231) by using the value ranges of concepts as 'names' of those concepts, and quantification over truth values in lieu of quantification over (Gödel numbers of) statements, thereby sidestepping the need for Gödel numbers or analogous coding devices.

Given the diagonalization result above, we can immediately generate the *Liar paradox*. Applying Theorem 1 to our falsity predicate results in a sentence Λ such that:

$$\Lambda = (\Lambda = \sim(\forall y)(y = y))$$

is a theorem. But this straightforwardly entails:

$$\sim(\forall y)(y = y)$$

a patent contradiction.

Thus, we can prove an analogue of Gödel's diagonalization lemma within the *Grundgesetze* and use it to construct the *Liar paradox*. The reader might wonder why we have made so much of these results, however. After all, we already knew that the *Grundgesetze* (including BLV) was inconsistent, so the news that one can construct the *Liar paradox* as well as *Russell's paradox* within Frege's system is not exactly earth-shattering news (although the 'naturalness' of the construction of the *Liar paradox* in the *Grundgesetze* is somewhat surprising, at least to the author).

The interest of these results lies in their connection to Frege's attempted fix of the *Grundgesetze* in the appendix to Volume II, to which we now turn.

2 Diagonalization and the Appendix to Grundgesetze

A quick examination of Theorem 1 reveals that the full strength of BLV is not required in order to prove the full, biconditional form of diagonalization. Instead, we merely need the resources to infer line (2):

$$\Phi(F(\S F))$$

from line (5):

$$(\exists Z)(\S F = \S Z \wedge Z(\S F) = Z(\S F) \wedge \Phi(Z(\S F))$$

In order to get from (5) to (2), we do not need it to be the case that concepts with the same value range are *always* co-extensive. Instead, we merely need concepts which have the same value range to agree on their shared value-range. Thus, we can recapture Theorem 1 by replacing BLV with the (prima facie weaker) Fixed-Point Principle for value-ranges:

$$\text{FPP: } (\forall X)(\forall Y)(\S(X) = \S(Y) \rightarrow (X(\S X) = Y(\S X)))$$

If FPP holds, then we can move from:

$$\S F = \S Z$$

to:

$$F(\S F) = Z(\S F)$$

and thus from:

$$\Phi(Z(\S F))$$

to:

$$\Phi(\text{F}(\S\text{F}))$$

The moral is simply this: Any principle meant to replace BLV and provide identity conditions for value ranges cannot, on pain of *Liar*-induced contradiction, imply FPP. Surprisingly, Frege was aware of this moral: In response to the detection of *Russell's paradox*, and without any (apparent) knowledge that the *Liar paradox* could also be derived within the *Grundgesetze*, Frege isolated FPP as exactly the problematic consequence of BLV.

In the appendix of Volume II of the *Grundgesetze*, Frege begins his discussion of *Russell's paradox* by distinguishing between the two 'directions' of BLV:

$$\text{BLVa: } (\forall X)(\forall Y)((\forall z)(X(z) = Y(z)) \rightarrow \S X = \S Y)$$
$$\text{BLVb: } (\forall X)(\forall Y)(\S X = \S Y \rightarrow (\forall z)(X(z) = Y(z)))$$

He notes that, if we are to individuate concepts extensionally (an assumption he is unwilling to give up), then BLVa cannot be the problem – after all, *any* function *f* from concepts to objects will satisfy:

$$(\forall X)(\forall Y)((\forall z)(X(z) = Y(z)) \rightarrow fX = fY)$$

So BLVb must be where the problem lies, and Frege sets out to discover exactly what goes wrong with this principle. He outlines his strategy as follows:

> We shall now try to complete our inquiry by reaching the falsity of (Vb) as the final result of a deduction, instead of starting from (Vb) and thus running into a contradiction. ([1893], p.288 in the Frege Reader)

Frege thought that the best strategy for gaining an understanding of exactly what it is about BLVb that causes the problem is to begin by finding a simple, direct proof of its negation – relying on a *reductio* of BLVb via Russell's construction might be enough to show us that the principle is problematic, but it does little to show us *why* it is problematic.[2]

In other words, Frege requires a direct proof of:

$$(\exists X)(\exists Y)(\S X = \S Y \wedge (\exists z)(X(z) \wedge \neg Y(z)))$$

In searching for such a proof, Frege discovers that he can obtain a stronger result, which I have come to call:

> **Frege's Little Theorem**: For any function *f* from concepts to objects one can prove:
>
> $$(\exists X)(\exists Y)(f(X) = f(Y) \wedge X(f(X)) \wedge \neg Y(f(X)))$$

Informally, this is just the claim that, given any function from concepts to objects, there exist two concepts such that the function maps both concepts to the same object, yet the concepts differ on that very object.

Here is the rub: The instance of *Frege's Little Theorem* obtained by substituting the value range operator "§" for "*f*" is the negation of FPP! In other words, the principle that Frege identifies as the root of *Russell's*

[2] Frege's proto-constructivist rejection of *reductio* proofs as uninformative, at least in this particular case, deserves an essay all its own!

paradox is exactly the principle that is needed to prove the version of the diagonalization lemma given above.

The proof of *Frege's Little Theorem* is as follows (see Frege [1893], pp.285–288 in the *Frege Reader*, for Frege's original proof):

Proof: Given a function f from concepts to objects, let:

$$R(x) = (\exists Y)(x = f(Y) \wedge \neg Y(x))$$

Then:

(1)	$\neg R(f(R))$	Assump
(2)	$\neg(\exists Y)(f(R) = f(Y) \wedge \neg Y(f(R)))$	(1), Df. of R
(3)	$(\forall Y)(f(R) = f(Y) \rightarrow Y(f(R)))$	(2)
(4)	$R(f(R))$	(3)
(5)	$R(f(R))$	(1)–(4), *Reductio*
(6)	$(\exists Y)(f(R) = f(Y) \wedge \neg Y(f(R)))$	(5), Df. of R
(7)	$(\exists Y)(f(R) = f(Y) \wedge R(f(R)) \wedge \neg Y(f(R)))$	(5), (6)
(8)	$(\exists X)(\exists Y)(f(X) = f(Y) \wedge X(f(X)) \wedge \neg Y(f(X)))$	(7), Logic

Frege concludes that such 'fixed points' are the root of Russell's paradox:

> We can see that the exceptional case is constituted by the extension itself, in that it falls under only one of the two concepts whose extension it is; and we see that the occurrence of this exception in no way can be avoided. Accordingly the following suggests itself as the criterion for equality in extension: The extension of one concept coincides with that of another when every object that falls under the first concept, except the extension of the first concept, falls under the extension of the second concept likewise, and when every object that falls under the second concept, except the extension of the second concept, falls under the first concept likewise. ([1893], p.288 in *The Frege Reader*)

As a result, Frege suggests a modification of BLV:

$$\text{BLV*}\ (\forall X)(\forall Y)((\S X = \S Y) = (\forall z)((z \neq \S X \wedge z \neq \S Y) \rightarrow (X(z) = Y(z))))$$

According to the amended principle two concepts receive the same value range if and only if they hold of exactly the same objects other than their value ranges.

The inadequacy of Frege's BLV* is well-known, although the reasons commonly given for its failure are mistaken. The well-known works

addressing the technical aspects of BLV*, Frege's so-called 'way out', such as Quine [1955] and Geach [1956], report that Frege's amended principle is consistent, but inadequate for his purposes (or any substantial purpose, really), since it implies that at most one object exists. What they fail to appreciate, however, is that since Frege's *Grundgesetze* allows for quantification into sentential position, one can (without any version of BLV, amended or not) prove the existence of at least two objects (the true and the false).[3] In other words:

$$(\exists x)(\exists y)(x \neq y)^{4}$$

is a theorem of the *Grundgesetze*. As a result, from the perspective of Frege's *Grundgesetze*, BLV* is just as inconsistent as was BLV.

Proving that BLV* implies the existence of no more than one object (along the lines of Quine [1955] or Geach [1956], perhaps) and then noting the existence, within the *Grundgesetze*, of at least two truth values is certainly enough to show that BLV* is inconsistent. Such a demonstration does not, however, provide much insight into the particular reasons underlying this inconsistency, or into the roots of paradox more generally. Fortunately, it turns out that we can do better – we can derive a version of the diagonalization lemma directly using BLV* (although the proof turns out to be much more difficult than it was in the case of BLV!) and thus reconstruct the *Liar paradox* in the amended system. The remainder of this essay is devoted to doing just that, and hopefully arriving at some insight into the nature of paradoxes (both semantic and set-theoretic) along the way. Before providing the main proof, however, a number of lemmata are required, and we now turn to this task.

[3] Landini [2006] comes closest to this result,, since he proves that BLV* is inconsistent if the truth values are their own singletons, as Frege intended, and also proves that BLV* is inconsistent if the truth values are not value-ranges at all.

[4] This theorem can be simply obtained from two applications of existential introduction to any instance of: $\Phi \neq \sim \Phi$.

3 LEMMATA[5]

In this section we shall prove some preliminary results that will be needed in what follows. First, for convenience we shall introduce an abbreviation for the analogue of set-theoretic singletons within Frege's *Grundgesetze*:

$$\{b\} =_{df} \S(x = b)$$

In other words, the Fregean singleton of an object b is the extension of the concept that holds of exactly the objects that are identical to b – that is, the concept that holds of b and b alone. With this notation in place, we can prove our three required lemmata:

Lemma 1: BLV* $\Rightarrow (\forall x)(\forall y)((x = \{x\} \wedge y = \{y\}) \rightarrow x = y)$

Proof:

(1)	$a = \{a\} \wedge b = \{b\}$	Assump.
(2)	$(\forall z)((z \neq \{a\} \wedge z \neq \{b\}) \rightarrow ((z = \{a\}) = (z = \{b\})))$	Tautology
(3)	$(\forall z)((z \neq \{a\} \wedge z \neq \{b\}) \rightarrow ((z = a) = (z = b)))$	(1), (2)
(4)	$\{a\} = \{b\}$	(3), BLV*
(5)	$a = b$	(1), (4)
(6)	$(a = \{a\} \wedge b = \{b\}) \rightarrow a = b$	(1)–(5), CP
(7)	$(\forall x)(\forall y)((x = \{x\} \wedge y = \{y\}) \rightarrow x = y)$	(6)

Thus, for any two objects, if they are both identical to their own singletons, then they are identical to each other. We shall call this result L1 in what follows.

Lemma 2: BLV* $\Rightarrow (\forall x)(\forall y)(\{x\} = \{y\} \rightarrow x = y)$

Proof:

(1)	$\{a\} = \{b\}$	Assump.
(2)	$(\forall z)((z \neq \{a\} \wedge z \neq \{b\}) \rightarrow ((z = a) = (z = b)))$	(1), BLV*
(3)	$(a \neq \{a\} \wedge a \neq \{b\}) \rightarrow ((a = a) = (a = b))$	(2)
(4)	$(a \neq \{a\} \rightarrow a = b)$	(1), (3)

[5] This section is deeply indepted to the excellent discussion of these issues in Burgess [1998] and [2005].

(5)	$(b \neq \{a\} \wedge b \neq \{b\}) \rightarrow ((b = a) = (b = b))$	(2)
(6)	$(b \neq \{b\} \rightarrow a = b)$	(1), (5)
(7)	$a \neq b$	Assump.
(8)	$a = \{a\} \wedge b = \{b\}$	(4), (6), (7)
(9)	$a = b$	(8), L1
(10)	$a \neq b \rightarrow a = b$	(7)–(9), CP
(11)	$a = b$	(10)
(12)	$\{a\} = \{b\} \rightarrow a = b$	(1)–(11), CP
(13)	$(\forall x)(\forall y)(\{x\} = \{y\} \rightarrow x = y)$	(12)

Thus, for any two objects, if the singletons of those objects are identical, then the objects are themselves identical. We shall call this L2 in what follows.

Lemma 3: BLV* $\Rightarrow (\forall x)(x = \{\{x\}\} \rightarrow x = \{x\})$

Proof:

(1)	$a = \{\{a\}\}$	Assumpt
(2)	$(\forall z)((z \neq \{a\} \wedge z \neq \{\{a\}\}) \rightarrow ((z = \{\{a\}\} = (z = \{a\}))$	Tautology
(3)	$(\forall z)((z \neq \{a\} \wedge z \neq \{\{a\}\}) \rightarrow ((z = a) = (z = \{a\}))$	(1), (2)
(4)	$\{a\} = \{\{a\}\}$	(3), BLV*
(5)	$a = \{a\}$	(4), L2
(6)	$a = \{\{a\}\} \rightarrow a = \{a\}$	(1)–(5), CP
(7)	$(\forall x)(x = \{\{x\}\} \rightarrow x = \{x\})$	(6)

Thus, any object that is identical to the singleton of its singleton is also identical to its singleton. We shall call this L3 in what follows.

The basic idea underlying these three results is that the singletons provided by BLV* are extremely well-behaved[6], and, in addition, for the most part they behave much as one would expect (e.g. Lemma 2 is a straightforward consequence of the extensionality axiom found in most standard set theories). We will make extensive use of this 'good' behavior in our revised proof of the diagonalization lemma.

[6] Lemmas 1 and 3 are both theorems of the non-well-founded set theories AFA and FAFA, but neither is a theorem of the non-well-founded set theory known as BAFA (see Aczel [1988] for details).

4 BLV* AND THE DIAGONALIZATION LEMMA

Before proving the diagonalization lemma, we need a new diagonalization relation:

$$\text{Diag*}(x, y) = (\exists Z)(y = \{\S Z\} \wedge x = Z(y))$$

"Diag*" holds between x and y if and only if y is the singleton of the value-range of some concept Z and x is the truth value obtained by applying Z to y – that is, of applying Z to the singleton of the value-range of Z. We can now prove our new version of diagonalization:

> **Theorem 2**: In the *Grundgesetz–BLV+BLV**, for any predicate $\Phi(x)$, there is a sentence G such that:
>
> $$\Phi(G) = G$$
>
> is a theorem.

Proof: Given $\Phi(x)$, let:

$$F(y) = (\exists x)(\text{Diag*}(x, y) \wedge \Phi(x))$$
$$G = F(\{\S F\})$$

($\rightarrow$)

(1)	$\Phi(G)$	Assump.
(2)	$\Phi(F(\{\S F\}))$	Df. of G
(3)	$(\forall x)(F(x) = F(x)) \wedge F(\{\S F\}) = F(\{\S F\}) \wedge \Phi(F(\{\S F\}))$	(2)
(4)	$(\exists Z)((\forall x)(F(x) = Z(x)) \wedge Z(\{\S F\}) = Z(\{\S F\}) \wedge \Phi(Z(\S F)))$	(3)
(5)	$(\exists Z)(\S F = \S Z \wedge Z(\{\S F\}) = Z(\{\S F\}) \wedge \Phi(Z(\{\S F\}))$	(4), BLV*
(6)	$(\exists Z)(\{\S F\} = \{\S Z\} \wedge Z(\{\S F\}) = Z(\{\S F\}) \wedge \Phi(Z(\{\S F\}))$	(5), BLV*
(7)	$(\exists x)(\exists Z)(\{\S F\} = \{\S Z\} \wedge x = Z(\{\S F\}) \wedge \Phi(x))$	(6)
(8)	$F(\{\S F\})$	(7), Df.of F
(9)	G	(8), Df.of G

(←)

(1)	G	Assump.
(2)	$F(\{\S F\})$	(2), Df. of G
(3)	$(\exists x)(\exists Z)(\{\S F\} = \{\S Z\} \wedge x = Z(\{\S F\}) \wedge \Phi(x))$	(3), Df. of F
(4)	$(\exists Z)(\{\S F\} = \{\S Z\} \wedge Z(\{\S F\}) = Z(\{\S F\}) \wedge \Phi(Z(\{\S F\})))$	(4)
(5)	$\{\S F\} = \{\S R\} \wedge R(\{\S F\}) = R(\{\S F\}) \wedge \Phi(R(\{\S F\}))$	Assump.
(6)	$\{\S F\} = \{\S R\} \wedge \Phi(R(\{\S F\}))$	(5)
(7)	$\S F = \S R \wedge \Phi(R(\{\S F\}))$	(6), L2
(8)	$\sim\Phi(F(\{\S F\}))$	Assump.
(9)	$F(\{\S F\}) \neq R(\{\S F\})$	(7), (8)
(10)	$\{\S F\} = \S F$	(7), (9), BLV*
(11)	$(\forall z)((z \neq \{\S F\} \wedge z \neq \S F) \rightarrow((z = \S F) = F(z)))$	(10), BLV*
(12)	$(\forall z)(z \neq \S F \rightarrow \sim F(z)))$	(11)
(13)	$a \neq \S F$	Assump.
(14)	$\{\{a\}\} \neq \S F$	(10), (13), L2
(15)	$\sim F(\{\{a\}\})$	(12), (14)
(16)	$\sim(\exists x)(\exists Z)(\{\{a\}\} = \{\S Z\} \wedge x = Z(\{\{a\}\}) \wedge \Phi(x))$	(15), Df. of F
(17)	$(\forall x)(\forall Z)((\{\{a\}\} = \{\S Z\} \wedge x = Z(\{\{a\}\})) \rightarrow \sim\Phi(x))$	(16)
(18)	$(\forall Z)(\{\{a\}\} = \{\S Z\} \rightarrow \sim\Phi(Z(\{\{a\}\})))$	(17)
(19)	$\{\{a\}\} = \{\{a\}\} \rightarrow \sim\Phi(a = \{\{a\}\})$	(18)
(20)	$\sim\Phi(a = \{\{a\}\})$	(19)
(21)	$a = \{\{a\}\}$	(2), (7), (8), (9)[7]
(22)	$a = \{a\}$	(21), L3
(23)	$a = \S F$	(22), L1
(24)	$a \neq \S F \rightarrow a = \S F$	(13)–(23), CP
(25)	$a = \S F$	(24)
(26)	$(\forall x)(x = \S F)$	(25)
(27)	$F(\{\S F\}) = \S F$	(26)
(28)	$R(\{\S F\}) = \S F$	(26)
(29)	$F(\{\S F\}) = R(\{\S F\})$	(27), (28)
(30)	$\Phi(F(\{\S F\}))$	(7), (29)

[7] A note on the passage from line (20) to line (21) is probably appropriate, since at first glance it might look like a non-sequitur. The reasoning, in more detail, is as follows: The expression "$F(\{\S F\})$" names the true (line (2)), and the truth values named by "$F(\{\S F\})$" and "$R(\{\S F\})$" are distinct (line (9)), so (since "$R(\{\S F\})$" must name a truth value) "$R(\{\S F\})$" names the false. So, Φ holds of the false (line (7)), but not the true (line 8). So, since "$a = \{\{a\}\}$" must name a truth value, and Φ fails to hold of it, it must name the true. Although tedious, this line of reasoning could of course be explicitly reconstructed within the *Grundgesetze* itself.

(31) $(\sim\Phi(F(\{\S F\}))) \rightarrow \Phi(F(\{\S F\}))$ (8)–(30), CP
(32) $\Phi(F(\{\S F\}))$ (31)
(33) $\Phi(F(\{\S F\}))$ (5)–(32), EE
(34) $\Phi(G)$ (33), Df. of G

We can now apply this new version of the diagonalization result to the falsity predicate defined in 1 above to obtain a contradiction, much as before. Thus, contrary to the well-established folk-wisdom, Frege's amended version of the *Grundgesetze*, as presented in the appendix to the second volume of this work, is no more consistent than the *Russell* and *Liar paradox* prone system presented in the first volume.

4 What Has Gone Wrong?

Thus, Frege's amended version of the *Grundgesetze* obtained by replacing BLV with BLV* is inconsistent. But we might fairly ask what went wrong. After all, doesn't *Frege's Little Theorem* provide a deep insight into the working of abstraction principles – one that suggests that BLV* was on the right track?

The correct answer to this question is that *Frege's Little Theorem* provides some insight into the roots of these paradoxes, but not enough.. As a result, the restriction on extensions suggested by *Frege's Little Theorem* – that is, the replacement of BLV with BLV* – does not go far enough. In order to see this we need merely note that a much stronger version of *Frege's Little Theorem* can be proven:

> **Generalized Frege's Little Theorem**: A function f is extension-injective on @ iff:
>
> $$(\forall X)(\forall Y)(f(@X) = f(@Y) \rightarrow @X = @Y)$$
>
> Given any extension-injective function f and any abstraction operator @:
>
> $$(\exists X)(\exists Y)(@X = @Y \wedge X(f(@X)) \wedge \neg Y(f(@X)))$$

In order to prove this result, we need a generalization of the Russell predicate (i.e. a unique Russellesque predicate for each function f):

f-Russell Predicate: $R_f(x) =_{df} (\exists Y)(x = f(@Y) \wedge \neg Y(x))$

Proof:

(1)	$\neg R_f(f(@R_f))$	Assump.
(2)	$\neg(\exists Y)(f(@R_f) = f(@Y) \wedge \neg Y(f(@R_f)))$	(1), df. of R_f
(3)	$(\forall Y)(f(@R_f) = f(@Y) \rightarrow Y(f(@R_f)))$	(2)
(4)	$f(@R_f) = f(@R_f) \rightarrow R_f(f(@R_f))$	(3)
(5)	$R_f(f(@R_f))$	(4)
(6)	$\neg R_f(f(@R_f)) \rightarrow R_f(f(@R_f))$	(1)–(5), CP
(7)	$R_f(f(@R_f))$	(6)
(8)	$(\exists Y)(f(@R_f) = f(@Y) \wedge \neg Y(f(@R_f))$	(7), df. of R_f
(9)	$f(@R_f) = f(@P) \wedge \neg P(f(@R_f)$	Assump.
(10)	$@R_f = @P \wedge \neg P(f(@R_f)$	(9)
(11)	$@R_f = @P \wedge R_f(f(@R_f)) \wedge \neg P(f(@R_f)$	(7), (10)
(12)	$(\exists X)(\exists Y)(@X = @Y \wedge X(f(@X)) \wedge \neg Y(f(@X)))$	(11)
(13)	$(\exists X)(\exists Y)(@X = @Y \wedge X(f(@X)) \wedge \neg Y(f(@X)))$	(8)–(12), EE

We can now see what went wrong in Frege's attempted 'fix' of the *Grundgesetze*: If we assume BLV*, then it follows that the singleton operator is an extension-injective function (since it is injective in general, according to Lemma 2 above). Given that the singleton operator is extension-injective, however, *Generalized Frege's Little Theorem* tells us that there must be two concepts that have the same extension yet which disagree on the *singleton* of that extension. It is this fact that is not accounted for by BLV*, which only allows for concepts which have the same extension but which disagree on that extension (not its singleton), and it is for this reason that BLV* fares no better than its predecessor.[8]

[8] A careful examination of the proof of diagonalization above will convince the reader that *Generalized Frege's Little Theorem* is exactly what is at issue. In effect, what the proof of diagonalization does is to find two concepts that have the same extension, but which might differ on the singleton of that extension. If they do, however, then the singleton of this extension must be identical to the extension itself, and as a result our three lemmata can be utilized to obtain the desired result.

5 LESSONS LEARNED

The ultimate failure of Frege's attempt to salvage his life's work does not imply that the it contains nothing of value. I will conclude by identifying two lessons that can, and should, be drawn from all of this.

The first is that we should take care in attributing the inadequacies of BLV* to some sort of panicked, half-hearted attempt by Frege to amend his formal system. Quine describes this common attitude to the appendix of *Grundgesetze Volume II*:

> It is scarcely to Frege's discredit that the explicitly speculative appendix now under discussion, written against time in a crisis, should turn out to possess less scientific value than biographical interest. Over the past half century the piece has perhaps had dozens of sympathetic readers who, after a certain amount of tinkering, have dismissed it as the wrong guess of a man in a hurry. (1955, p.152)

While the 'fix' might have been written in a hurry, and BLV* is inconsistent, the discussion leading up to it has much to teach us about the mathematics of abstraction principles in general and the roots of *Russell's Paradox* (unsurprisingly) and the *Liar Paradox* (surprisingly) in particular. In this respect, *Frege's Little Theorem*, and the amended version of BLV based on this result, are not the incorrect guesses of a man in a hurry, but on the contrary represent a deep insight into the puzzling nature of abstraction and the paradoxes that can arise from its unfettered application. Unfortunately, this insight was not enough to save the amended system from similar paradoxes, but this does not mean that Frege's Little Theorem was not an insight nonetheless.

This brings us to the second lesson. Connections are often drawn between the *Liar paradox* and *Russell's paradox* (and between the semantic and set-theoretic paradoxes more generally), but these connections tend to be quite loose, relying on the intuition that circularity of some vicious sort is at the root of both phenomena[9] (for a project that draws the connections much more tightly, however, the reader is urged to consult Cook [2007]!). The construction of the *Liar paradox* within Frege's system, and his identification of the exact principle that is the root of both this paradox and the one communicated to him by Russell, suggests that further study of Frege's system (or modern variants that retain object-level quantification into sen-

[9] For a careful examination of the arguments for and against such a connection, see Terzian [2008].

tential position, such as that provided in Landini [2006]) hold promise for a deeper understanding of these paradoxes individually and of the links that might or might not bind them together.[10]

REFERENCES

Aczel, P. [1988], *Non-Well-Founded Sets*, Stanford: CSLI.

Beall, J. (ed.) [2007], *Revenge of the Liar*, Oxford: Oxford University Press.

Boolos, G. J. Burgess, & R. Jeffrey, [1989] *Computability & Logic*, 5th Ed., Cambridge: Cambridge University Press.

Burgess, J. [1998], "On a Consistent Subsystem of Frege's *Grundgesetze*", *Notre Dame Journal of Formal Logic 39*: pp.274–278.

Burgess, J. [2005], *Fixing Frege*, Princeton: Princeton University Press.

Cook, R. [2007] "Embracing Revenge: On the Indefinite Extensibility of Language", in Beall [2007]: pp.31–52.

Frege, G. [1893], 1903 *Grundgezetze der Arithmetik I & II*, Hildesheim: Olms.

Frege, G. [1997] *The Frege Reader*, M. Beaney (ed.), Oxford: Blackwell.

Geach, P. [1956] "On Frege's Way Out", *Mind 65*, pp.408–409.

Gödel, K. [1992] *,* New York: Dover.

Hieke, A. & H. Leitgeb (eds.) [2008], *Reduction and Elimination in Philosophy and the Sciences: Preproceedings of the 31st International Wittgenstein Symposium*, Kirchberg am Wechsel, Austrian Ludwig Wittgenstein Society.

Landini, G. [2006] "The Ins and Outs of Frege's Way Out", *Philosophia Mathematica 14*, pp.1–25.

Quine, W.V.O. [1955] "On Frege's Way Out", *Mind 64*, pp.145–159.

Shapiro, S. [1991], *Foundations Without Foundationalism: The Case for Second-Order Logic*, Oxford: Oxford University Press.

Terzian, G. [2008], "Structure of the Paradoxes, Structure of the Theories: A Logical Comparison of Set Theory and Semantics", in Hieke & Leitgeb [2008]: pp.347–349.

[10] Thanks go to Philip Ebert, Marcus Rossberg, Greg Taylor, and Giulia Terzian, for helpful discussion of earlier versions of this paper and suggestions for improvement. In addition, the present paper has benefited greatly from feedback received at *Arché: The Philosophical Research Centre for Logic, Language, Metaphysics, and Epistemology at the University of St Andrews* and to the *31st International Wittgenstein Symposium*, where earlier versions of this paper were presented.