The Liar Paradox in Frege's Set Theory

Floris T. van Vugt
University College Utrecht
Utrecht University, The Netherlands

May 2, 2005

Abstract

This paper is meant to provide an exposition of the nature of Russell's Paradox that is meant to make clear that it is a formalised version of the principle that is expressed in the *Liar Paradox* — the uttering of the phrase "This, what I say now, is not true." First of all the paradoxical nature of this statement will be scrutinised. The discussion of Russell's Paradox will be preceded by a discussion of Frege's definition of logical inference and his proposition for the foundations of mathematics — in particular his concept of a *Werthverlauf* that was meant to bring functions into the realm of analysis of objects. Consequently, it will be explained why Russell's Paradox is crucial to Frege's undertaking. Subsequently, using elementary Model Theory in logic there will be an explanation of the modern resolution of the paradox in Zermelo–Fraenkel Set Theory. Finally, there will be a discussion of the fundamental characteristic of self–reference that is exhibited in Russell's Paradox and made impossible in the theory of Zermelo–Fraenkel.

1 Introduction

The many paradoxes that exist in both mathematical and everyday thinking whose fascination has led to believe that the many similarities might point at a common source to at least some of them.

In this context the present paper will constitute a modest argument for the thesis that the Liar Paradox and Russell's Paradox rely on the same deeper principle.

2 Preliminaries

It will be assumed that the reader is familiar with the concept of the natural numbers, which is the smallest set such that 0 is in the set and if any number is in the set, its successor is so too, i.e. the numbers $0, 1, \ldots$.

The notation $n \in \mathbb{N}$ will mean only that n is a natural number.

3 The Liar Paradox

The paradox consists in an utterance, which is the following:

The Liar Paradox $LP \doteq$ "That, which I say now, is false.".

The paradoxical nature of this statement is exhibited in the failure to answer the question whether the utterance is true or false, since if it is true, then it must be so that that which was said is false, which is a contradiction, but if it were false, then that which is said is false, which means that the utterance is true, which contradicts that it was false. Both options lead to a contradiction.

First of all it should be noted here that this paradox is only a problem if one's aim is to predicate either truth or falsity of all statements, since this paradox gives an example of a statement to which both its truth and falsity lead to a contradiction. If only one of the two would lead to a contradiction, as in the case of the utterance "This, what I now say, is true," where the hypothesis that it is false is contradictory, which is not problematic since the opposite is not.

Formally, this predication of truth or falsity to each proposition can be seen as a function $\mathcal{T}$ that associates with each element of the set of all propositions $\mathcal{P}$ an element from the set $\{\top, \bot\}$, where we think of $\top$ and $\bot$ as representing truth and falsity, but, since they are at this stage introduced merely syntactically, for that matter, they could be thought of as meaning any thing. Among other things, it is evident that we would want that for all propositions of the form "A is false," (where A is again a proposition) it would be that $\mathcal{T}$ associates truth with it only if it associates falsity with A, and falsity with it if truth with A.

$$\begin{aligned} \mathcal{T}(\text{“}A \text{ is false”}) = \top \quad &\text{only if} \quad \mathcal{T}(A) = \bot \\ \mathcal{T}(\text{“}A \text{ is false”}) = \bot \quad &\text{only if} \quad \mathcal{T}(A) = \top \end{aligned}$$

The existence of LP shows that such a function $\mathcal{T}$ cannot exist, since LP is of the form "A is false," where A is LP.

$$\begin{aligned} \mathcal{T}(LP) = \top \quad &\text{only if} \quad \mathcal{T}(LP) = \bot \\ \mathcal{T}(LP) = \bot \quad &\text{only if} \quad \mathcal{T}(LP) = \top \end{aligned}$$

It should be clear at this point that this is an equivalent, yet slightly more involved way of putting it. The formal language will be justified later on.

An endeavour of this sort, however, was, to its fullest extent, attempted only by Leibniz. Frege and later Hilbert defined their increasingly modest programs as finding a predication of this sort for the whole of mathematics respectively for all axiomatic subsystems. Nevertheless their efforts were equally unfruitful.

Secondly, in view of our purposes later on there will be made a distinction between the following kinds of reasons for a proposition to be true: (1) *formal*, i.e. when the proposition or its negation as a whole is a premise, (2) *intrinsic*, i.e. when that which is expressed by it is true.

The paradoxical nature can then be made clear in the following scheme of reasoning:

$$LP \text{ is true} \overset{\text{intrinsic}}{\rightarrow} \text{The uttered proposition is false} \overset{\text{formal}}{\rightarrow} LP \text{ is false}$$
$$LP \text{ is false} \overset{\text{intrinsic}}{\rightarrow} \text{The uttered proposition is true} \overset{\text{formal}}{\rightarrow} LP \text{ is true}$$

It also becomes clear that the paradox relies on the problem that we cannot distinguish, either intrinsically or formally, between LP and "LP is false."

Since what has been paradoxical was that neither truth nor falsity could be predicated of LP, the formal importance of this paradox vanishes if one allows the function $\mathcal{T}$ to leave certain propositions undecided, i.e. that there exists $p \in \mathcal{P}$ such that $\mathcal{T}(p) \neq \top$ and $\mathcal{T}(p) \neq \bot$.

That this cannot be so is in this case the *Law of the excluded middle.* It can be brought up, however, that if one would extend the function $\mathcal{T}$ such that it now associates to each proposition not only truth or falsity, but also possibly a “neutral truth-value”: $\{\top, \bot, 0\}$. Indeed this particular paradox is then solved, it would simply put $\mathcal{T}(LP) = 0$. But if we then extend LP to $LP' \doteq$“This proposition is neither true nor of neutral truth value.” then it cannot be that $\mathcal{T}(LP') = 0$, since it would obviously be false. It shows that the problem is more fundamental than the absence of a third truth–value. In fact, this argument can be repeated if one would allow another truth–value for propositions that are neither true, false or neutral — a being undecided even on this higher level. The problem lies thus in the fact that the proposition LP or any variant of it, would appear in the domain $\mathcal{P}$ of $\mathcal{T}$.

4 Frege’s Foundations of Mathematics

4.1 Outline

Frege attempted to provide an axiomatisation for Set Theory, which is a theory of fundamental status since the objects it defines — sets, functions, relations — are taken for granted in other branches of mathematics.

At the basis of his theory lies the concept of a generic *object* [1]. In addition to the obvious choice of numbers and sets, Frege saw that the abstraction of his theory allowed for any objects in the broad sense of the word to be considered objects in his language, for instance chairs, tables or men. *Functions* are then introduced as wholly different entities that associate certain objects with others. The habitual functions that associate numbers to numbers are then incorporated, since numbers are examples of objects, but for instance a price list can also be treated as one, associating with each product an amount of money, which both are objects. Each function has a natural number associated to it, referred to as its “arity,” which is the number of arguments it takes. For example, some price lists may have their prices depend on a qualities of the customer, for instance a train tickets’ price may depend on your age, and if that is the only criterion it is a function of arity 1. In number theory *addition* is an example of a function of arity 2 that associates with every two objects a and b their sum $a + b$.

It is important to note in this context that, in our and Frege’s concept of it, a function that takes particular objects as its arguments associates one and only one object with them. If we write $f(a_0, \ldots, a_n) = b$ we mean that the function f associates with the definite objects referred to by $a_0, \ldots, a_n$

the object referred to as b. This distinction can be thought of as a formal example of the distinction between *sense* and *reference*[3], when the name "$f(a_0, \ldots, a_n)$" has a different meaning — the object that f associates with the $n+1$–tuple $(a_0, \ldots, a_n)$ — from "b," — the object called "b" — but the same reference. The symbol "=" thus equates reference and not meaning. A more elaborate discussion of this matter falls outside the domain of this paper.

4.2 Singular Sentences

Frege tries to formalise these concepts into a language that is defined as follows.

Objects can be referred to by names agreed upon in advance. For instance "2" can be taken to refer to the natural number which is the successor of the successor of "0." The name "King Lear" refers to a, probably, imaginary literary character, which, too, can be considered an object.

Functions are referred to by unique names. If f is the name of a function of arity n, $f(a_0, \ldots, a_n)$ is a short way of referring to the object that f associates with the collection of $n+1$ definite objects referred to as $a_0, \ldots, a_n$.

Terms are elements of the smallest set that contains all object names — such as "3" or "π", but also "The present king of France" [3] — and for each function f of any arity $n \in \mathbb{N}$, if $a_0, \ldots, a_n$ are terms, then $f(a_0, \ldots, a_n)$ is a term. For instance, since 2 and 3 are terms, also $2+3$ is a term, since it is obtained by applying the addition function with 2 and 3 as arguments.

Sentences are terms that are either true or false.

In particular, it is clear that if s and t are terms, $s = t$ is a sentence, since it can be either true or false. For Frege, equivalence is a 2–ary function $=(a_0, a_1)$ that associates with all pairs of objects a_0 and a_1 truth if they refer to the same object, and falsity if they do not.

It is also clear that if s is a sentence, the utterance "it is not true that s" is a sentence, and it is true if s is false and false if s is true. This phrase is written as $\neg s$[1]. Furthermore, if s and t are sentences, "s and t are both true," is also a sentence, and it is true if and only if s is true and t is true. This sentence is written as $s \wedge t$. Likewise, "s or t" is written as $s \vee t$, which is true if and only if s and t are not both not true. The sentence "in this case, the condition holds that if s is true, then t is true" is written as $s \rightarrow t$ and is called *material implication.* It is true if and only if it is not true that s is true but not t. In this way it is equivalent to $\neg(s \wedge \neg t)$.[2]

Concepts are sentences in which one of the terms is replaced by a placeholder.

A placeholder is often referred to as a variable — to distinguish it from object names whose references are thought of as fixed — and its appearance in sentence turns it into a function that associates with each substitution of an object name for the placeholder the truth value truth or falsity. This point is crucial and an example will make it clear. "2," "3" and "5" are object names. Addition is a function f (though more often the symbol $+$ is used, in this discussion such a choice is arbitrary) such that for all a and b, $f(a, b) = a + b$. "$f(2, 3) = 5$" is a sentence and it is true, contrary to "$f(2, 2) = 5$." If one uses x as a symbol for a place–holder, "$f(2, x) = 5$" is a *concept*, which can be expressed as "to be such that it yields 5 if 2 is added to it." Concepts are conventionally given capital letters as their names. Thus the mentioned concept can be written as $F(x)$ and it associates falsity with all x except for $x = 3$.

4.3 Quantified Statements

Frege then defines quantifiers to allow singular statements to be applied to collections of objects. "For all objects, in what follows referred to as x, it holds that $M(x)$," where $M(x)$ is a concept, is written formally as $\forall x(M(x))$.

[1]This is a more modern notation and it differs from Frege's. Since formulas of each way of writing can without exception be translated into one another, the current notation is chosen for reasons of typesetting convenience and compatibility with other works.

[2]To understand that it must be so, it is often pedagogical to think of $s \rightarrow t$ as a condition on a situation. For instance, one could be concerned with a collection of marbles. s is the proposition that there is at least a blue marble. t is the condition that there is at least a red marble. Obviously the condition "If there is at least one blue marble, there must also be at least one red marble" is satisfied, thus true, in all cases unless there is at least one blue marble but none red.

Likewise, "There is at least one object, in what follows referred to as x, such that $M(x)$," is written as $\exists x(M(x))$. Notice that there is at least one object x for which $M(x)$ is true, if and only if, it is not true that for all objects it is not true that $M(x)$. In this way $\exists x(M(x))$ is true if and only if $\neg\forall x(\neg M(x))$. This implies that one could write all statements in which $\forall$ or $\exists$ appear in terms of only one of these quantifiers.

More complex statements can now formally be written. For instance, if $M(x)$ is the concept "to be mortal", and $P(x)$ is the concept "to be a person," then "All persons are mortal" is expressed by $\forall x(P(x) \rightarrow M(x))$. Likewise, the statement "All and only persons are mortal" is written as $\forall x\,[(P(x) \rightarrow M(x)) \wedge (M(x) \rightarrow P(x))]$, or, if one uses the convention to abbreviate the statement $(a \rightarrow b) \wedge (b \rightarrow a)$ to $a \leftrightarrow b$, one writes $\forall x(P(x) \leftrightarrow M(x))$.

4.4 Axioms

Frege felt that there were formulas true and rules of inference valid simply by virtue of their *form*. In these cases there was no need for *intrinsic* reasoning and therefore these formulas and these forms of reasoning could be formulated as axioms and rules of inference, respectively, in his *Begriffssschrift*.

Tautologies are formulated as schema's — which means that for any of the variables $a, b, \ldots$ that occur in them, any proposition can be substituted, for example $\neg a$ or $\neg a \implies b$ — and they are the following, for any term t:[3]

Ax0. $a \rightarrow \neg\neg a$

Ax1. $a \rightarrow a$

Ax2. $a \rightarrow (b \rightarrow a)$

Ax3. $(a \rightarrow (b \rightarrow c)) \rightarrow ((a \rightarrow b) \rightarrow (a \rightarrow c))$

Ax4. $(\neg b \rightarrow \neg a) \rightarrow (a \rightarrow b)$.

Ax5. $\forall x(A(x)) \rightarrow A(t)$

[3]The axioms are formulated in terms of the syntactic connectives $\rightarrow$ and $\neg$ only. All the other connectives can be defined in terms of them. $a \vee b$ is equivalent to $\neg a \rightarrow b$ and $a \wedge b$ is equivalent to $\neg(\neg a \vee \neg b)$. Translation to the rudimentary formulation in terms of $\neg, \rightarrow$ is, however, unnecessarily tedious for more complicated sentences

The rules of inference are the following. They are written in the form $a, b \Rightarrow c$ to mean that from the sentences a and b one can infer the sentence c. This concept is made more formal later on.

MP. **Modus ponens** $a, a \rightarrow b \Rightarrow b$

AI. $a \rightarrow B(x) \Rightarrow a \rightarrow \forall x(B(x))$.

Frege's aim was that on the basis of these axioms and the rules of inference, all other sentences that we think of as tautologies could be derived. If successful, it frees the way for the definition of tautologies as those and only those derivable sentences and the axiomatisation would therewith be complete.

4.5 Proofs

Frege then defines that a proof of a particular sentence p is a finite succession of sentences such that

1. all sentences in the collection are
   (a) explicitly expressed premises, or
   (b) tautologies (Ax0.–Ax5.) given in the previous subsection, or
   (c) follow from earlier sentences in the collection by means of one of the two rules of inference given in the previous subsection as well, and
2. the sentence p is in the succession.

In modern days one writes $T \vdash a$ if there exists a proof of the proposition a whose hypotheses are included in T. If the proof does not include any hypotheses, instead of $\emptyset \vdash a$ (where $\emptyset$ is the set that contains no elements) one also writes $\vdash a$.

Frege's aim is here to define the notion of a proof in such a way that $T \vdash a$ if and only if p is true if all of the sentences in T are. That this is, in the end, so is not altogether trivial, but can be proven.

For instance, it is natural to think that if we say $M(a)$, then there exists at least an x such that $M(x)$. Formally, there should be a rule

$$\textbf{Existential Introduction} \qquad M(a) \Rightarrow \exists x(M(x)) \tag{1}$$

Indeed there is, and the following is its proof. It is written as a succession of numbered sentences, accompanied by an explanation in the right hand column of why the condition 1. mentioned above is satisfied in the case of that particular sentence.

1.	$M(x)$	premise
2.	$M(x) \rightarrow \neg\neg M(x)$	Ax0
3.	$\neg\neg M(x)$	1,2,MP
4.	$\neg\neg M(x) \rightarrow \neg\forall x(\neg M(x))$	Ax4, Ax5
5.	$\neg\forall x(\neg M(x))$	3,4,MP
6.	$\exists x(M(x))$	5, definition of $\exists$

It is important to notice that we have only proven it for the arbitrary but particular concept M and object a. The fact that we have not put any constraint whatsoever on it, allows us, reasoning outside the system, to assert that it can be repeated for any concept M and any object x. Moreover, the fact that it is a rule is saying that any quasi–proof in which $\exists x(M(x))$ is inferred from $M(a)$ can be expanded to form a full valid proof according to the criteria set down. This is the full extent of what is expressed by saying that equation 1.

Thus one can write $\{M(a)\} \vdash \exists x(M(x))$, where $\{M(a)\}$ denotes the set that contains, for that matter, at least the formula $M(a)$.

4.6 Course–of–values

Frege's functions have so far been wholly detached from the realm of objects. One could say that mathematically hardly anything is known about them — a relation as fundamental as equality has only been defined on the level of objects. In order for Frege to treat them in some way as objects he introduces the idea of *course–of–values* (*Werthverlauf* in German), and it is written as $\acute{\varepsilon} f(\varepsilon)$. A course–of–value, as opposed to a function, is thought of as an object.

The need for elaboration on the concept of a function developed until now becomes apparent from the example of the following functions:

$$\begin{aligned} f(x) &= x(x+1) \\ g(x) &= x^2 + x \end{aligned}$$

It is intuitively clear that they are the same functions, since they agree for each value for the argument x which object to associate with it, $\forall x(f(x) =$

$g(x)$). This fact can be easily proven within algebra. However, it has not been set down that this is a (1) condition, and (2) that this condition is sufficient for courses–of–values to be considered equivalent. The formalisation of this apparent triviality becomes embodied in Frege's *Law V*[4]:

$$\acute{\varepsilon} f(\varepsilon) = \acute{\alpha} g(\alpha) \leftrightarrow \forall x(f(x) = g(x)) \tag{2}$$

The formulation of this law has made it possible to express the equality of the functions f and g above.

Intuitively, the course–of–values can be seen as the collection of ordered couples that records all values that the function associates with all possible arguments. For instance, for the function $f(x) = x + 1$ it notes that 2 is $f(3)$, and $f(5)$ is 6. It should be stressed however, that this is an intuition that has guided the formulation of equation 2 but which should be kept out of consideration now that this latter has been formally set down.

For those functions that are concepts, the course–of–values therefore registers which objects are mapped to the truth, and which to falsity. In this way, it defines the collection of objects for which the condition that is expressed by them holds, which is then called the *extension* of the concept G, and it is denoted by εG.

Frege then went on to define what it means for an object to be a member of an extension. Intuitively, an object x is a member of an extension of a concept G if it is true that $G(x)$:

$$x \in \varepsilon G \doteq G(x)$$

Likewise, membership among objects can be defined as:

$$\begin{aligned} x \in y &\doteq \exists G(y = \varepsilon G \wedge x \in \varepsilon G) \\ x \in y &\doteq \exists G(y = \varepsilon G \wedge G(x) \end{aligned} \tag{3}$$

With the tools that have been developed so far, it can be proven that for

[4]Here the notation = for the equality among courses–of–values is chosen deliberately to emphasise the difference with the symbol = that is used for object comparison only.

any concept F, it holds that $\{F(x)\} \vdash x \in \varepsilon F$.[5]

$$\begin{array}{lcl} 1. & F(x) & \text{premise} \\ 2. & \varepsilon F = \varepsilon F & \text{eq2} \\ 3. & \varepsilon F = \varepsilon F \wedge F(x) & 1,2 \\ 4. & \exists G(\varepsilon F = \varepsilon G \wedge G(x) & \text{3, eq1} \\ 5. & x \in \varepsilon F & \text{4, eq3} \end{array}$$

Similarly, one can prove that each concept has an extension, a fact that later on will be crucial — one notices that if F is a concept, $F(x) = F(y)$ is the same as $F(y) \leftrightarrow F(x)$ — :

$$\begin{array}{lcl} 1. & \forall F(\varepsilon F = \varepsilon F \leftrightarrow \forall x(F(x) \leftrightarrow F(x))) & \text{eq2} \\ 2. & \forall F \forall x(F(x) \leftrightarrow F(x)) & \text{tautology} \\ 3. & \forall F(\varepsilon F = \varepsilon F) & \text{1,2} \\ 4. & \forall F \exists x(x = \varepsilon F) & \text{3,eq1} \end{array} \tag{4}$$

5 Russell's Paradox

5.1 Formal Derivation

Russell studied the system that Frege had started and found it possible to derive a contradiction. The presentation of the reasoning is based on that of Zalta[5].

In this section the contradiction will be proven on the basis of the *Law of Extensions*

$$\textbf{Law of Extensions} \qquad \vdash \forall F \forall x(x \in \varepsilon F \leftrightarrow F(x)) \tag{5}$$

[5] This proof uses $\{a, b\} \vdash a \wedge b$, which can be proven starting from the axioms given in the previous section. The notation eq2 refers to equation number 2, namely the *Law V*

Its proof is as follows[6]

$$
\begin{array}{lcl}
a1. & x \in \varepsilon F & \text{premise} \\
a2. & \exists G(\varepsilon G = \varepsilon F \wedge G(x)) & \text{eq3} \\
a3. & \exists G(\forall x(G(x) \leftrightarrow F(x)) \wedge G(x)) & \text{eq2} \\
a4. & [\forall x(G(x) \leftrightarrow F(x)) \wedge G(x)] \rightarrow F(x) & \text{tautology} \\
a5. & \exists G(F(x)) & \text{a3,a4} \\
a6. & F(x) & \text{a5} \\
a7. & \forall F \forall x(x \in \varepsilon F \rightarrow F(x)) & \text{a5,generalisation}
\end{array}
$$

$$
\begin{array}{lcl}
b1. & F(x) & \text{premise} \\
b2. & \varepsilon F = \varepsilon F & \text{eq4, identity} \\
b3. & \varepsilon F = \varepsilon F \wedge F(x) & \text{b1,b2} \\
b4. & \exists G(\varepsilon G = \varepsilon F \wedge G(x)) & \text{b3,eq1} \\
b5. & x \in \varepsilon F & \text{b5,eq3} \\
b6. & \forall F \forall x(x \in \varepsilon F \leftarrow F(x)) & \text{b6,generalisation}
\end{array}
$$

$$LE. \quad \forall F \forall x(x \in \varepsilon F \leftrightarrow F(x)) \quad \text{a7,b6} \tag{6}$$

On the basis of the *Law of Extensions* one can derive the *Naive Comprehension Axiom*[7], $\vdash \forall F \exists y \forall x(x \in y \leftrightarrow F(x))$, simply

$$
\begin{array}{lll}
1. & \forall F \forall x(x \in \varepsilon F \leftrightarrow F(x)) & \text{Law of Extensions} \\
2. & \forall F \exists y \forall x(x \in y \leftrightarrow F(x)) & \text{1,eq1}
\end{array} \tag{7}
$$

This point is crucial. We have now found that for all predicates there exists an object, whose elements are those objects for which the predicate is true.

[6]This proof proceeds by reasoning outside the formal system and in that sense is not strictly a proof according to the constraints that Frege defined. Since, however, writing out the proof in its full detail is complex and little illuminating, it is chosen to present an argument that shows clearly that a strict proof can be given.

[7]An interesting observation is that the quantifiers $\forall, \exists$ have been used not only to denote objects, but also more generally concepts. The language Frege has been using is a *Second Order Language*. Modern logic distinguishes the two and does, in ordinary set theory, not allow this ambiguity.

If we then take the predicate $F(x) \doteq \neg(x \in x)$ (which is the predicate that is true of all objects that are not elements of themselves), we find

$$\begin{array}{rcl} 3. & \exists y \forall x(x \in y \leftrightarrow \neg(x \in x)) & 2 \\ 4. & \forall x(x \in a \leftrightarrow \neg(x \in x)) & \text{3,take any } y \\ RP. & (a \in a) \leftrightarrow \neg(a \in a) & \text{4,take } a \text{ for } x \end{array}$$

Which shows the contradiction, where the enunciation $(a \in a)$ has taken the place of LP, since it can neither be true nor false.

6 Analysis

6.1 Importance

From a contradiction any proposition can be derived.

To show this, let us call the proposition that is contradicted a, such that by hypothesis $T \vdash a$, $T \vdash \neg a$. We show that there exists a rule that can be formulated as follows — where p is any proposition:

$$\textbf{Contradiction} \qquad a, \neg a \Rightarrow p \tag{8}$$

$$\begin{array}{rcl} 1. & a & \text{hypothesis} \\ 2. & \neg a & \text{hypothesis} \\ 3. & \neg p \to a & \text{1,Ax2} \\ 4. & \neg p \to \neg a & \text{2,Ax2} \\ 5. & (\neg p \to \neg a) \to ((\neg p \to a) \to p) & \text{Ax} \\ 6. & (\neg p \to a) \to p & \text{4,5,MP} \\ 7. & p & \text{3,6,MP} \end{array}$$

From the exposition made so far it becomes apparent that this problem is relevant, since (1) the concepts of objects, function, concept and extension are fundamental in mathematical analysis, and (2) the essence of *Law V* presents itself to us as tautological and therefore abandoning it would imply failure to provide an axiomatisation of our concepts mentioned before, which was precisely what Frege was aiming at.

Frege writes[2]

> [Russell's] discovery of the contradiction caused me the greatest surprise and, I would almost say, consternation, since it has shaken the basis on which I intended to build arithmetic. [...] It is all the more serious since, with the loss of my Rule V, not only the foundations of my arithmetic, but also the sole possible foundations of arithmetic, seem to vanish.

6.2 The predicate $F(x)\dot{=}\neg(x \in x)$

The predicate $F(x)\dot{=}\neg(x \in x)$ seems strange. It has, however, been well defined, since each of the symbols that it uses — $\neg, \in$ — are well defined, and it can be viewed as the conceptualisation of the negation of the sentence $a \in a$, which is a sentence since it is a term that is either true or false, by the definition of $\in$. It becomes more clear that there is a need to incorporate sentences of this sort from the fact that one can talk about this predicate F. For instance, the collection of all books is not again a book, thus it is true that F("the collection of all books"), but the list that records all lists is again a list, thus not F("list of all lists"). These examples, though necessarily artificial, show that the concept expressed by F is not meaningless.

Similarly, LP is, in any natural language, a grammatically valid sentence.

6.3 Nature of the Paradox

As mentioned before, LP does not lead to a contradiction unless one proceeds from the assumption that to each sentence one should attribute either truth or falsity. Russell writes[4]:

> Let w be the predicate: to be a predicate that cannot be predicated of itself. [...] [W]e must conclude that w is not a predicate.

Frege qualifies this by arguing that it is not meaningful to say that a predicate is predicated of any predicate — therefore also not of itself — and that it is only meaningful to talk of predication of objects. He acknowledges however that the proof of the paradox did not rely on this[2]:

> Incidentally, it seems to me that the expression "a predicate is predicated of itself" is not exact. A predicate is as a rule a first–level function, and this function requires an object as argument and cannot have itself as argument (subject). Therefore I would prefer to say "a notion is predicated of its own extension."

There is a correspondence between Russell's Paradox and the Liar Paradox. Following the definition of the predicate $F(x) \doteq \neg(x \in x)$, one calls the extension of this concept — which exists due to the Naive Comprehension Axiom — εF. Now the question is whether F can be predicated of its own extension, i.e. whether $F(\varepsilon F)$ or not. If $F(\varepsilon F)$ then, by the definition of F, it must be that $\neg(\varepsilon F \in \varepsilon F)$. Formally, however, this implies, by definition of $\in$, that $\neg F(\varepsilon F)$. Similarly for the contrary. From the scheme below the resemblance with LP becomes clear.

$$F(\varepsilon F) \stackrel{\text{intrinsic}}{\rightarrow} \neg(\varepsilon F \in \varepsilon F) \stackrel{\text{formal}}{\rightarrow} \neg F(\varepsilon F)$$
$$\neg F(\varepsilon F) \stackrel{\text{intrinsic}}{\rightarrow} \varepsilon F \in \varepsilon F \stackrel{\text{formal}}{\rightarrow} F(\varepsilon F)$$

7 Modern Set Theory

7.1 Model Theory

Model Theory tries to generalise and formalise mathematical derivation.

Language A language $\mathbb{L}$ consists of a set of *constants*, *function symbols* and *relation symbols*. To each function and relation symbol one associates a natural number n referred to as its arity as before.

Examples of constants are 0, 1, examples of functions are $+$ (addition) and $\cdot$ (for multiplication), both of arity 2, and an example of a relation symbol is $\leq$. It is important to note that at this stage these symbols are considered to be without meaning, which will be attributed to them later. In the following the examples will be taken starting from the language $\mathbb{L} \doteq \{0, 1, +, \cdot, \leq\}$, which is a language used to define the natural numbers.

Terms The terms of a language $\mathbb{L}$ are defined as the smallest set $\mathbb{T}_{\mathbb{L}}$ such that:

- any constant of $\mathbb{L}$ is a term
- any variable is a term
- if $t_1, \ldots, t_n$ are terms and f is an n–ary function, $f(t_1, \ldots, t_n)$ is a term.

Examples of terms are 0, a (where a is a variable – a symbol that does not appear in the language), but also $+(1,0)$, if $+$ is a 2–ary function.

Formulas The formulas of a language $\mathbb{L}$ are defined as the smallest set $\mathbb{F}_{\mathbb{L}}$ such that

- if s and t are terms, $s = t$ is a formula.
- if $t_1, \ldots, t_n$ are terms and R is a n–ary relation, $R(t_1, \ldots, t_n)$ is a formula.
- $\perp$ is a formula.
- if ϕ and ψ are formulas, then $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \to \psi$ and $\neg\phi$ are formulas.
- if ϕ is a formula and x is a variable, then $\forall x\phi$ and $\exists x\phi$ are formulas.

Examples of formulas are $0 \leq 1$ or $a \leq 0 \to a \leq 1$. Often one refers to a formula ϕ in which there appears a variable a that does not appear as part of a quantifier ($\forall a$ or $\exists a$) as $\phi(a)$.

Theory A theory of a language $\mathbb{L}$ is a set of formulas $\mathbb{F}_{\mathbb{L}}$.

Model A model for a language $\mathbb{L}$ is a set $M_{\mathbb{L}}$ such that

- With each constant c in $\mathbb{L}$ there is associated an element $c^{M_{\mathbb{L}}} \in M_{\mathbb{L}}$, called the *interpretation* of the constant c.
- With each n–ary function f in $\mathbb{L}$, there is associated a function $f^{M_{\mathbb{L}}}$, called its *interpretation*, that associates with each set of n elements of $M_{\mathbb{L}}$ an element of $M_{\mathbb{L}}$.
- With each n–ary relation R in $\mathbb{L}$, there is associated a set $R^{M_{\mathbb{L}}}$, called its *interpretation*, of n–tuples of elements of $M_{\mathbb{L}}$.

7.2 Tarski's Semantics

Interpretation of terms If $M_{\mathbb{L}}$ is a model of the language $\mathbb{L}$, then one extrapolates the association with each term t of an element $t^{M_{\mathbb{L}}}$ to terms that are complexes in which no variables occur.

- if t is a constant, by definition there has already been associated an element $t^{M_{\mathbb{L}}}$.
- if t is of the form $f(t_1, \ldots, t_n)$ then $t^{M_{\mathbb{L}}}$ is the element that is associated by the function inside the model, $f^{M_{\mathbb{L}}}(t_1^{M_{\mathbb{L}}}, \ldots, t_n^{M_{\mathbb{L}}})$.

Sufficiency A model $M_{\mathbb{L}}$ is said to be a model of a formula $\phi_{\mathbb{L}}$ — one writes $M_{\mathbb{L}} \models \phi_{\mathbb{L}}$ — if, on the basis of the form of the formula,

- If $\phi_{\mathbb{L}}$ is $\perp$, it never holds that $M_{\mathbb{L}} \models \phi_{\mathbb{L}}$.
- If $\phi_{\mathbb{L}}$ is $s = t$, where s and t are terms, it holds if $s^{M_{\mathbb{L}}} = t^{M_{\mathbb{L}}}$
- If $\phi_{\mathbb{L}}$ is $R(t_0, \ldots, t_n)$, it holds if the collection of $(t_0^{M_{\mathbb{L}}}, \ldots, t_n^{M_{\mathbb{L}}})$ is in the set $R^{M_{\mathbb{L}}}$.
- If $\phi_{\mathbb{L}}$ is $\phi_1 \wedge \phi_2$ or $\phi_1 \vee \phi_2$ or $\phi_1 \rightarrow \phi_2$ or $\neg\phi_1$, the expression $M_{\mathbb{L}} \models \phi_{\mathbb{L}}$ holds according to the truth-values of the constituents as defined before.
- If $\phi_{\mathbb{L}}$ is of the form $\forall x \phi_1(x)$ then $M_{\mathbb{L}} \models \phi_{\mathbb{L}}$ holds if, substituting a new constant m for x, for all possible interpretations $m^{M_{\mathbb{L}}} \in M_{\mathbb{L}}$, $M_{\mathbb{L}} \models \phi_1(m)$.
- If $\phi_{\mathbb{L}}$ is of the form $\exists x \phi_1(x)$ then $M_{\mathbb{L}} \models \phi_{\mathbb{L}}$ holds if, substituting a new constant m for x, for at least one interpretation $m^{M_{\mathbb{L}}} \in M_{\mathbb{L}}$, $M_{\mathbb{L}} \models \phi_1(m)$.

Model of a Theory A model is a model *of a theory* $\mathbb{T}_{\mathbb{L}}$ if $M_{\mathbb{L}} \models \phi$ for each $\phi \in \mathbb{T}_{\mathbb{L}}$. One writes $M_{\mathbb{L}} \models \mathbb{T}_{\mathbb{L}}$.

An example of a simple theory in the elementary language $\mathbb{L} \doteq \{a, b\}$ where a and b are constants, is the theory that consists only in the formula $a = b$. Any set is a model of this theory, as long as the constants a and b are interpreted as the same element in the set (which is a sufficient condition for the formula $a = b$ to be true), i.e. $a^{M_{\mathbb{L}}} = b^{M_{\mathbb{L}}}$. Notice that there might be any number of elements in the set. If one would extend the theory to $\{a = b, \forall x(x = a \vee x = b)\}$, the number of elements is restricted to 1, since due to the latter axiom all elements in the set should be equal either to the interpretation of a or b, and due to the former these interpretations are identical.

7.3 Zermelo–Fraenkel Set Theory

Russell's Paradox has shown that there is a need for a stricter set theory (as opposed to the *Naive Set Theory* as it was promoted by Frege and others) that does not allow the definition of a set of all sets that are not an element of themselves. Nowadays the most widely used set theory is that of Zermelo–Fraenkel, which is a theory in the language which consists only in the relation symbol for membership $\in$ — i.e. $\mathbb{L} = \{\in\}$ and there are no functions nor constants — and its axioms are the formulas from $\mathbb{F}_{\mathbb{L}}$ listed below — where one writes $\overrightarrow{z}$ for $z_1, \ldots, z_n$.

$$
\begin{array}{rc}
\text{Ext.} & \forall x_0 \forall x_1 (\forall y (y \in x_0 \leftrightarrow y \in x_1) \rightarrow x_0 = x_1 \\
\text{Pair} & \forall x_0 \forall x_1 \exists y \forall z (z \in y \leftrightarrow (z = x_0) \vee (z = x_1)) \\
\text{Union} & \forall x \exists y \forall z (z \in y \leftrightarrow \exists v (v \in x \wedge z \in v)) \\
\text{Powerset} & \forall x \exists y \forall z (z \in y \leftrightarrow \forall v (v \in z \rightarrow v \in x)) \\
\text{Infinity} & \exists w (\exists x \, [x \in w \wedge \neg \exists y (y \in x)] \wedge \\
 & \wedge \forall x \, [x \in w \rightarrow \exists y (y \in w \wedge \forall z (z \in y \leftrightarrow z \in x \vee z = x))]) \\
\text{Found.} & \forall x (\exists y (y \in x) \rightarrow \exists y \, [y \in x \wedge \neg \exists z (z \in x \wedge z \in y)]) \\
\text{Repl.} & \forall z_1 \ldots \forall z_n (\forall x \forall y_0 \forall y_1 \, [\phi(x, y_0, \overrightarrow{z}) \wedge \phi(x, y_1, \overrightarrow{z}) \rightarrow y_0 = y_1] \rightarrow \\
 & \rightarrow \forall y \exists w \forall y \, [y \in w \leftrightarrow \exists x (x \in y \wedge \phi(x, y, \overrightarrow{z}))]) \\
 & \text{for each } (n+2)\text{–ary } \{\in\}\text{–formula } \phi, (n \geq 0)
\end{array}
$$

The set of all these axioms is referred to as the theory ZF. One thinks of any model M_{ZF} such that $M_{ZF} \models ZF$ as the collection of all sets. If one picks one of these models, all sets are precisely the elements $x \in M_{ZF}$. In this way, sets are thought of as precisely the objects in M_{ZF} and thus the names become interchangeable.

It is important to notice that the last axiom is an axiom schema, it is a template and it is given that for each formula ϕ, there is the corresponding version of the *Replacement*–axiom. In fact, ZF has an infinity of axioms, since there is an infinity of possible formulas.

Using these axioms, one can prove that there are sets with particular properties, for instance, that there exists an empty set, $\emptyset \in M_{ZF}$, such that $\forall x \neg (x \in \emptyset)$[8]. This shows that not existing is not the same as being an empty

[8]This follows from the axiom "Repl." Taking $\phi(x, y, z_1) = (y \in y) \wedge \neg (y \in y)$ — a formula that is never true — for any set x it follows that there exists a set w that has the properties of the empty set, namely that it has no elements.

set. The counterparts of extensions in Frege's theory are referred to as *classes*. A class for a formula $\phi(x)$ is a collection of sets such that the sets and only the sets that satisfy the formula $\phi(x)$ belong to the class, and it is referred to as $\{x|\phi(x)\}$. However — contrary to Frege's Naive Set Theory, where all concepts have an extension – not every class $\{x|\phi(x)\}$ is a set (i.e. there are formulas ϕ such that it does not hold that $ZF \models \exists y \forall z (z \in y \leftrightarrow \phi(z))$. In other words, the *Naive Comprehension Axiom* that was derivable in Frege's theory, is negated by any counterexample, such as will be given below). The most important example of this is the class of all sets that are not elements of themselves.

7.4 Solution to Russell's Paradox

In ZF one can claim that there is, in any model M of ZF, not necessarily a set corresponding to the class $\{x|\neg(x \in x)\}$, since its existence cannot be proven directly — as in the set theory proposed by Frege — and a contradiction can be derived from its hypothetical existence.

This is done as follows. Assume that $R \doteq \{x|\neg(x \in x)\}$. Thus for all x one can write $x \in R \leftrightarrow \neg(x \in x)$. If it were so that R is a set, then one can take $x = R$ and find $R \in R \leftrightarrow \neg(R \in R)$, which is a contradiction. This shows that the assumption that R is a set is false.

This shows that RP cannot be formulated in ZF and thus there is no more contradiction. The price has been a theory that is less intuitive, since of each class one would have to prove that it is a set for it to be treated according to the axioms.

It is interesting to note that this argument has a counterpart in the set theory that was proposed by Frege. The conclusion is essentially the same contradiction $R \in R \leftrightarrow \neg(R \in R)$. The difference is that in the former set theory there was already a proof that this set R existed. In ZF this proof is absent and thus the contradiction counts as a disproof[9].

The contradiction therefore finds its root in the *Naive Comprehension Axiom*, equation 7, which is a direct corollary of the *Law V*. The Naive Comprehension Axiom can be viewed as performing a similar function as the desire to predicate either truth or falsity of all propositions, which

[9]The assumption here is that ZF is consistent — i.e. that no contradiction can be derived from the axioms —. This is a hidden assumption that — as Gödel's Incompleteness Theorems have it — cannot be proven nor disproven, since those proofs should start from ZF.

Finally, Tarski's semantics and the definition of the meaning of the quantifiers have prevented any theory from expressing second–order formulas, since all quantifiers range over elements of the model, which, in the case of ZF, are only the sets and not functions or relations.

8 Conclusion

In conclusion, it has been argued in this paper that the problems around Frege's Naive Set Theory arise from the fact that all concepts have an extension, which, in the framework of ZF, would mean that to all classes there corresponds a set such that the two sets assigned to the classes are equal if and only if the formulas defining the classes are equivalent. Precisely this cannot be held. It is essentially the self–referencing nature that would make the one–to–one–correspondence of classes and sets problematic.

References

[1] Gottlob Frege. Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In Jean van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, 1967. Harvard University Press.

[2] Gottlob Frege. Letter to russell. In Jean van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, 1967. Harvard University Press.

[3] Bertrand Russell. On denoting. *Mind, New Series*, 14(56):479–493, 1905.

[4] Bertrand Russell. Letter to frege. In Jean van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, 1967. Harvard University Press.

[5] Edward N. Zalta. Frege's logic, theorem, and foundations for arithmetic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2005.