1

# On a Russellian paradox about propositions and truth

# Andrea Cantini

**Abstract.** We deal with a paradox involving the relations between propositions and sets (appendix B of *Principles of Mathematics*), and the problem of its formalization. We first propose two (mutually incompatible) abstract theories of propositions and truth. The systems are predicatively inspired and are shown consistent by constructing suitable inductive models.

We then consider a reconstruction of a theory of truth in the context of (a consistent fragment of) Quine's set theory. The theory is motivated by an alternative route to the solution of the Russellian difficulty and yields an impredicative semantical system, where there exists a high degree of self-reference and yet paradoxes are blocked by restrictions to the diagonalization mechanism.

# 1. A paradox of Russell concerning the type of propositions

The doctrine of types is put forward tentatively in the second appendix of the *Principles* (§500). It is assumed that to each propositional function  $\phi$  is associated a range of significance, i.e. a class of objects to which the given  $\phi$  applies in order to produce a proposition; moreover, precisely the ranges of significance form types. However there are objects that are not ranges of significance; these are just the individuals and they form the lowest type. The next type consists of classes or ranges of individuals; then one has classes of classes of objects of the lowest type, and so on.

Once the hierarchy is accepted, new difficulties arise; in particular, if one accepts that *propositions form a type* (as they are the only objects of which it can be meaningfully asserted that they are true or false). This is a crucial point and leads Russell to a contradiction, which explicitly involves semantical notions.

<sup>\*</sup>Research supported by Università degli Studi di Firenze and MIUR. Part of this paper has been presented under the title *Relating KF to NF* at the Symposium *Russell 2001*, June 2-5 2001, München.

First of all, since it is possible to form types of propositions, there are more types of propositions than propositions, by Cantor's argument. But then there is an argument which, if sound, apparently refutes Cantor's theorem: we can inject types of propositions into propositions by appealing to the notion of logical product<sup>1</sup>.

Indeed, let m be a type of propositions; to m we can associate a proposition  $\Pi m$  which expresses that "every proposition of m is true" (to be regarded as a possibly infinitary conjunction or logical product). Now, if m and n are (extensionally) different types, the propositions  $\Pi m$  and  $\Pi n$  must be regarded as distinct for Russell, i.e. the map  $m \mapsto \Pi m$  is injective. Of course, if one were to adopt the extensional point of view, and hence equivalent propositions should be identified, no contradiction could be derived. But for Russell nobody will identify two propositions if they are simply logically equivalent; the proper equality on propositions must be much more fine-grained than logical equivalence. For instance, the proposition "every proposition which is either an m or asserts that every element of m is true, is true" is not identical with the proposition "every element of m is true" and yet the two are certainly logically equivalent.

Of course, the conflict can easily be rephrased into the form of an explicit paradox: if we accept  $\{p \mid \exists m(\Pi m = p \land p \notin m)\} = R$  as a well-defined type<sup>2</sup>, we have, by injectivity of  $\Pi$ ,

$$\Pi R \in R \Leftrightarrow \Pi R \notin R,$$

whence a contradiction.

So, if we stick to injectivity of  $\Pi$ , we have to change some basic tenet, e.g. to reject the assumption that propositions form one type, and hence that they ought to have various types, while logical products ought to have propositions of the same type as factors.

This will be eventually the base of the 1908 solution, but here Russell reputes the suggestion as harsh and artificial; as the reader can verify from the text<sup>3</sup>, he still believes that the set of all propositions is a counterexample to Cantor's theorem.

The *Principles of Mathematics*, as its coeval Fregean second volume of the *Grundgesetze*, conclude with an unsolved antinomy and Russell declares that

<sup>&</sup>lt;sup>1</sup>A first discussion of the problem already appears in §349 of [21], where Russell deals with cases in which the conclusion of Cantor's theorem is plainly false. In this context, he mentions the crucial paradoxical embedding of classes into propositions, without solution: "I reluctantly leave the problem to the ingenuity of the reader", p.368.

<sup>&</sup>lt;sup>2</sup>Of course, in the definition of R the quantifier ranges over types of propositions.

<sup>&</sup>lt;sup>3</sup>See [21], p.527, footnote: ".It might be doubted whether the relation of propositions to their logical products is one-one or many one. For example, does the logical product of p and q and r differ from that of pq and r? A reference to the definition of the logical product (p.21) will set this doubt at rest; for the two logical products in question, though equivalent, are by non means identical. Consequently there is a one-one relation of all ranges of propositions to some propositions, which is directly contradictory to Cantor's theorem."

"what the complete solution of the difficulty may be", he has not succeeded in discovering; "but as it affects the very foundations of reasoning", he earnestly commends "the study of it to the attention of all students of logic." ([21], p.528)

# 1.1. A formal outlook

In the literature there have been attempts to connect Russell's contradiction on propositions to modal paradoxes (Oksanen [16]); quite recently, Cocchiarella ([5]) shows how to resolve the contradiction in intensional logics that are equiconsistent with NFU, Quine's set theory with atoms. There is also a book of P. Grim, which is entirely devoted to related issues ([11]).

In the following we first show that Russell's argument can naturally be formalized and resolved within the framework of a theory PT of operations, propositions and truth, which is closely related to Aczel's (classical) Frege structures ([1]). We then consider a variant of PT, proving that the very notion of propositional function defines a propositional function (this is refuted in PT).

PT will comprise the axioms of combinatory logic with extensionality (see [2]) and the abstract axioms for truth and propositions (T, P respectively). We assume that the language contains individual constants  $\dot{\rightarrow}$ ,  $\dot{\wedge}$ ,  $\dot{\neg}$ ,  $\dot{\forall}$ , representing the logical operations  $\rightarrow$ ,  $\wedge$ ,  $\neg$ ,  $\forall$  and individual constants  $\doteq$ ,  $\dot{T}$ ,  $\dot{P}$  representing the ground predicate symbols =, T and P. We implicitly assume suitable independence axioms among the dotted symbols (e.g.  $\dot{\neg} \neq \dot{\vee}$ ,  $\dot{\neg}x \neq \dot{\vee}x$ , etc.), granting provability of a formal analogue of the unique read-abilility property.

It is then straightforward to define an operation  $A \mapsto [A]$ , which assigns to each formula of the language a term [A] with the same free variables of A, which designates the "propositional object" associated to  $A^{4}$ . As usual, since we can define lambda abstraction, we can identify class abstraction  $\{x \mid A\}$ with  $\lambda x.[A]$ .

We also define:

$$\begin{array}{rcl} PF(f) &\Leftrightarrow & (\forall x)(P(fx)); \\ \Pi f &:= & \dot{\forall}(\lambda x(\dot{\rightarrow}(fx)x)); \\ a \equiv_e b &\Leftrightarrow & (Ta \leftrightarrow Tb); \\ a =_e b &\Leftrightarrow & \forall x(T(ax) \leftrightarrow T(bx)); \\ f \subseteq P &\Leftrightarrow & \forall x(T(fx) \rightarrow P(x)); \\ \dot{\lor}ab &:= & \dot{\neg}(\dot{\land}(\dot{\neg}a)(\dot{\neg}b)); \\ & \exists f &:= & \dot{\neg}(\dot{\forall}(\lambda x.\dot{\neg}(fx))). \end{array}$$

<sup>4</sup>This is an idea of D.Scott; see [3].

As to the terminology, if PF(f) is assumed, we say that f is a propositional function; sometimes we use  $x \in f$  instead of T(fx).

The point we wish to raise is that at the very beginning of his foundational work, Russell hits upon arguments which naturally require a framework where semantical notions as well as the logical notion of set (as extension of propositional function) live on the same par.

**Definition 1.** We list the basic principles for propositions and truth; we essentially extend the principles implicit in the definition of *Frege structure* à la Aczel (see [1]), with a few extra axioms.

- $\mathrm{P1} \ P([x=y]) \land .(T([x=y]) \leftrightarrow x=y);$
- P2  $T(a) \rightarrow P(a);$ P3  $P(a) \rightarrow T([P(a)]);$ P4  $P([P(a)]) \rightarrow P(a);$ P5  $P([T(a)]) \leftrightarrow P(a);$ P6  $T([T(a)]) \leftrightarrow T(a);$ P7  $P(a) \rightarrow (\neg T(a) \rightarrow T(\neg a));$ P8  $T(\neg a) \rightarrow \neg T(a);$ P9  $P(\neg a) \leftrightarrow P(a);$ P10  $P(a) \land (T(a) \to P(b)) \to P(\dot{\to}ab);$ P11  $P(\rightarrow ab) \rightarrow (T(a) \rightarrow P(b));$ P12  $P(\rightarrow ab) \rightarrow (T(a) \rightarrow T(b) \rightarrow T(\rightarrow ab));$ P13  $T(\rightarrow ab) \rightarrow (T(a) \rightarrow T(b));$ P14  $P(a) \wedge P(b) \leftrightarrow P(\dot{\wedge}ab);$ P15  $T(\dot{\wedge}ab) \leftrightarrow T(a) \wedge T(b);$ P16  $\forall x P(fx) \leftrightarrow P(\dot{\forall}f);$ P17  $T(\dot{\forall}f) \leftrightarrow \forall x(T(fx)).$

**Remark 1.** (i) The axioms above imply a *strict interpretation* of (classically defined) disjunction and existential quantifier. By contrast with Aczel's original framework, it is assumed that it makes sense to use the predicates 'to be

a proposition' and 'to be true' as logical contructors on the same par as the standard logical operators.

(ii) The question "to which kinds of objects truth can be rightly attributed" is a debatable issue among philosophers. For instance, in the Tarskian approach or in Kripke's paper [15] truth is attributed to (objects representing) *sentences*, i.e. to elements of an inductively defined syntactical category. On the contrary, we underline that here propositions form an abstract collection of objects, which are only required to meet certain broad closure conditions; being members of a combinatory structure, they can be freely combined for obtaining self-referential side effects for free. In general no induction on propositions is assumed (even if this might be true in some model, cf.§2). Also, the system is non-committal about the (delicate) question of defining a proper intensional equality for propositions. Here propositions inherit a neutral equality relation, determined by the applicative behaviour in the ground structure.

**Lemma 1** (PT). (i) If PF(f) and  $f \subseteq P$ , then  $\Pi f$  is a proposition such that

$$T(\Pi f) \leftrightarrow (\forall x (T(fx) \to T(x)))$$

Moreover:

$$\begin{array}{rccc} T(a) & \leftrightarrow & T(\dot{\neg}(\dot{\neg}a));\\ T([\neg T(a)]) & \leftrightarrow & T(\dot{\neg}a);\\ T([P(a)]) & \leftrightarrow & P(a);\\ & & & \exists x \neg T([\neg P(x)]);\\ P(\dot{\lor}ab) & \leftrightarrow & P(a) \land P(b);\\ P(a) \land P(b) \rightarrow (T(\dot{\lor}ab) & \leftrightarrow & T(a) \lor T(b));\\ P(\dot{\exists}f) & \leftrightarrow & \forall x P(fx);\\ \forall x P(fx) \rightarrow (T(\dot{\exists}f) & \leftrightarrow & \exists x T(fx)). \end{array}$$

(ii) We also have:

$$\Pi f = \Pi g \quad \to \quad f = g$$
$$PF(a) \land PF(b) \land a =_e b \quad \to \quad (\Pi a \equiv_e \Pi b)$$

(iii)  $\Pi$  is not extensionally injective, i.e. there exist propositional functions a, b such that

$$\Pi a \equiv_e \Pi b \land \neg (a =_e b).$$

*Proof.* As to (i), we only check the first claim (using axioms for  $\rightarrow$ ,  $\dot{\forall}$ )

$$\begin{aligned} PF(f) \wedge f \subseteq P &\Rightarrow P(fx) \wedge (T(fx) \to P(x)) \\ &\Rightarrow P(\to(fx)x) \\ &\Rightarrow PF(\lambda x. \to (fx)x) \\ &\Rightarrow P(\Pi f) \\ &\Rightarrow T(\Pi f) \leftrightarrow \forall x (T(fx) \to T(x)) \end{aligned}$$

(ii): apply injectivity of  $\dot{\forall}$ ,  $\rightarrow$  and extensionality for operations. (iii): choose  $a = \{[K = K], [S = S]\}$  and  $b = \{[K = K]\}$ ; then  $\Pi a \equiv_e \Pi b$ , but a and b are extensionally distinct propositional functions). QED.

Clearly, we can derive the Tarskian T-schema:

**Proposition 1.** If A is an arbitrary formula, then PT proves:

$$P([A]) \to (T([A]) \leftrightarrow A)$$

Proposition 2 (Russell's Appendix B, [21]). The term

$$\{x \mid \exists m(PF(m) \land m \subseteq P \land x \notin m \land x = \Pi m)\}$$

does not define a propositional function, provably in PT.

*Proof.* Let

$$R := \{ x \mid \exists m (PF(m) \land m \subseteq P \land x \notin m \land x = \Pi m) \}.$$

Assume by contradiction that PF(R). Then, by applying the closure conditions of P, T and lemma 1:

$$(\forall x \in R)(P(x)),$$

Hence:

$$P(\Pi R)$$
.

Now we have, for some m:

$$\Pi R \in R \Rightarrow \Pi R = \Pi m \wedge PF(m) \wedge m \subseteq P \wedge \Pi R \notin m.$$

By the previous lemma ( $\Pi$  is 1-1), R = m and hence  $\Pi R \notin R$ . But  $\Pi R \notin R$  implies  $\Pi R \in R$ , since R is a propositional function of propositions.

**Lemma 2** (PT). If P is a propositional function, then PF itself is a propositional function.

*Proof.* If P is a propositional function, we have:

$$\Rightarrow \quad \forall x.P([P(x)]) \Rightarrow \quad \forall x.P([P(fx)]) \Rightarrow \quad \forall fP([\forall x.P(fx)]) \Rightarrow \quad \forall fP([PF(f)])$$

**Theorem 1** (PT). *PF and P are not propositional functions.* 

*Proof.* By the previous lemma, it is enough to show that PF is not a propositional function.

Assume PF is a propositional function. The axioms on the relation between P and  $\forall$  imply:

$$\forall x \forall u. P([P(xu)]).$$

Hence using the axiom  $P([P(xu)]) \rightarrow P(xu)$ , we conclude:

 $\forall x. PF(x),$ 

against the previous proposition.

Alternative argument: if  $\lambda x.PF(x)$  is a propositional function, we can show, with the implication axioms that

$$W := \{ x \mid PF(x) \land (PF(x) \to \dot{\neg}(xx)) \}$$

is a propositional function. Then the standard Russell paradox arises.

The conclusion is that under relatively mild hypotheses (a notion of truth which obeys to classical laws, endowed with an abstract notion of proposition) the paradox disappears. Clearly, no propositional function reasonably defining the power class of the collection of propositions can exist in the above framework; on the same par, the collection of propositions cannot give rise to a well-defined propositional function. The solution is compatible with Russell's no-class theory: it could be assumed that the universe of classes exactly includes those collections which are represented by terms of the form  $\{x \mid A(x)\}$ , where A(x) defines a propositional function (cf. the model construction of §2).

# 1.2. An alternative theory of truth and propositions

**Definition 2.** We describe a variant AT of the system PT, which is so devised that the assumption "PF(x) is a proposition" is consistent:

A1 
$$P([x = y]) \land .(T([x = y]) \leftrightarrow x = y);$$

A2 
$$T(a) \rightarrow P(a)$$

A3  $\forall xT([\neg P(x)]);$ A4  $P([T(a)]) \leftrightarrow P(a);$ A5  $T([T(a)]) \leftrightarrow T(a);$ A6  $P(a) \rightarrow (\neg T(a) \rightarrow T(\neg a));$ A7  $T(\neg a) \rightarrow \neg T(a);$ A8  $P(\neg a) \leftrightarrow P(a);$ A9  $P(a) \land P(b) \leftrightarrow P(\land ab);$ A10  $T(\land ab) \leftrightarrow T(a) \land T(b);$ A11  $\forall xP(fx) \leftrightarrow P(\forall f);$ 

A12  $T(\dot{\forall}f) \leftrightarrow \forall x(T(fx)).$ 

The typical principle of the system is A3, according to which no claim about P(a) can be internally true. A3 implies with A2, A8 the following principle  $(\Lambda)$ :

 $\forall x.P([P(x)])$ 

It follows from  $(\Lambda)$  that AT is inconsistent with the axiom:

 $P([P(x)]) \to P(x).$ 

Indeed, let L be the Liar object  $L = \dot{\neg}(L)$ . Since P([P(L)]) holds, we have P(L) and we can conclude with the negation axioms that  $T(\dot{\neg}L) \leftrightarrow \neg T(L)$ , whence a contradiction.

Moreover, if we apply  $(\Lambda)$  and A11, we obtain:

**Proposition 3** (AT).  $\lambda x.PF(x)$  is a propositional function.

Observe also that A3 implies that the PT-axiom  $P(a) \to T([P(a)])$  is inconsistent with AT (choose a := [x = y]).

On the other hand, similarly to PT, we can prove in AT:

$$T([\neg T(a)]) \leftrightarrow T(\dot{\neg}a);$$

In order to appreciate the difference in strength between AT and PT, it may be of interest to compare the closure properties of propositional functions in either system. First of all, both systems are closed under *elementary comprehension*. Indeed, if  $\vec{a}$  is a finite list  $a_0, \ldots a_n$  of variables distinct from x, we say that  $A(x, \vec{a})$ , where  $FV(A) \subseteq \vec{a}, x$  is *elementary in*  $\vec{a}$ , if A is inductively generated from atomic formulas of the form  $t = s, t \in a_i$  (with  $0 \le i \le n$  and no variable of  $\vec{a}$  free in t, s) by means of  $\neg$ ,  $\land$  and quantification on variables not occurring in the list  $\vec{a}$ .

Let  $PT \cap AT$  be the theory consisting of those axioms, which are common to PT and AT.

**Proposition 4** (PT $\cap$ AT). If  $A(x, \vec{a})$  is elementary in  $\vec{a}$  and each  $a_i$  of the list  $\vec{a}$  is a propositional function, then  $\{x \mid A(x, \vec{a})\}$  is a propositional function such that

$$\forall u(u \in \{x \,|\, A(x, \vec{a})\} \leftrightarrow A(u, \vec{a}))$$

However, while propositional functions are closed under disjoint union (or join) in PT, the same cannot be proved in AT.

Let us justify this claim. We introduce a kind of sequential conjunction  $\odot$ , due to Aczel [1]:

$$a \odot b := a \land (a \to b)$$

Clearly PT proves:

$$P(a) \land (T(a) \to P(b)) \to P(a \odot b) \land T(a \odot b) \leftrightarrow T(a) \land T(b)$$

Since we have combinatory logic as underlying theory, we can assume to have an ordered pairing operation  $x, y \mapsto (x, y)$  with projections  $u \mapsto u_0, u \mapsto u_1$ . Therefore it makes sense to define

$$\Sigma(a, f) := \{ u \, | \, (u = (u_0, u_1) \land \, u_0 \in a \, \odot \, u_1 \in f(u)_0) \}$$

**Proposition 5.** (i) PT proves that, if a is a propositional function and fx is a propositional function whenever  $x \in a$ , then  $\Sigma(a, f)$  is a propositional function satisfying (J) :

$$u \in \Sigma(a, f) \leftrightarrow u = (u_0, u_1) \land u_0 \in a \land u_1 \in f(u)_0$$

(ii) AT proves the weak power class axiom: for every propositional function a, there exists a propositional function  $Pow^{-}(a)$  such that

$$\begin{aligned} \forall u(u \in Pow^{-}(a) & \to & PF(u) \land u \subseteq a); \\ \forall u(PF(u) \land u \subseteq a & \to & \exists b(PF(b) \land b \in Pow^{-}(a) \land u =_{e} b)) \end{aligned}$$

As to the proof, (i) is an immediate consequence of the definition of  $\odot$ , while (ii) essentially depends on the axiom that P(x) is a proposition for every x, once we set

$$Pow^{-}(a) = \{ u \mid PF(u) \land \exists b (PF(b) \land u = b \cap a) \}$$

**Corollary 1.** AT+(J) is inconsistent.

This follows by adapting Russell's paradox along the lines of Feferman [7].

### 1.3. PT and AT are non-cantorian

The claim means that in neither truth theory there is a good analogue of the power set in terms of propositional functions. The reason is given by the results below, which are provable in the common subtheory of PT and AT (so they do not involve properties of strong implication nor the fact that  $\lambda x.PF(x)$  is a propositional function).

**Definition 3.** (i) (A formula of our language)  $\varphi(x)$  is called *extensional* iff  $\forall f \forall g(PF(f) \land PF(g) \land f =_e g \land \varphi(f) \to \varphi(g))$ ;

(ii) a formula  $\varphi(x)$  such that  $\forall x(\varphi(x) \to PF(x))$  is non-trivial, provided there are propositional functions x, y such that  $\varphi(x), \neg \varphi(y)$ ).

**Lemma 3** ("Inseparability", in PT $\cap$ AT). Assume that  $\varphi_1$ ,  $\varphi_2$  are extensional formulas and there exist propositional functions  $x_1$ ,  $x_2$  such that  $\varphi_1(x_1)$ ,  $\varphi_2(x_2)$ . Then  $\varphi_1$  and  $\varphi_2$  are PF-inseparable, i.e. for no propositional function  $x_3$  we can have:

$$\forall u(PF(u) \land \varphi_2(u) \to u \notin x_3) \land \forall u(PF(u) \land \varphi_1(u) \to u \in x_3)$$

*Proof.* Assume that  $x_3$  is a propositional function such that

$$\forall u(PF(u) \land \varphi_2(u) \to u \notin x_3)$$

It is enough to produce a propositional function  $g := g(x_1, x_2, x_3)$  such that

 $g \notin x_3 \land \varphi_1(g).$ 

Choose by the fixed point lemma an element g such that g = Gg, where

$$Gh = \{ u \mid (u \in x_1 \land h \notin x_3) \lor (u \in x_2 \land h \in x_3) \}$$

Then, using the common axioms on  $\wedge$ ,  $\neg$  and the assumption that  $x_3$ ,  $x_2$ ,  $x_1$  are propositional functions, we have that g is a propositional function.

If  $g \in x_3$ , then  $g =_e x_2$ . Since  $\varphi_2(x_2)$ , also  $\varphi_2(g)$ ; thus  $g \notin x_3$ .

Hence  $g \notin x_3$ , which yields  $g =_e x_1$ . But  $\varphi_1(x_1)$ ; so  $\varphi_1(g)$  by extensionality. QED

**Theorem 2** (PT $\cap$ AT). No propositional function f can be both extensional and non-trivial.

So, for instance, there cannot exist a propositional function playing the role of the power set of  $\{\emptyset\}$ , i.e. whose range is exactly the collection of all propositional functions  $\subseteq \{\emptyset\}$ : for this would be non-trivial and extensional.

The results above are extensions to the context of Frege structures of results holding for theories of explicit mathematics (see [4]).

#### 2. About models

We outline two inductive model constructions for PT and AT (respectively).

# 2.1. PT-models

The basic idea is to consider any given combinatory algebra as ground universe and to produce, by generalized inductive definition, the collections of propositions and truths. This cannot be simply rephrased as a standard monotone inductive definition, because the clause for introducing a proposition of implicative form makes use (negatively) of the collection of truths. Nevertheless, we can adapt a trick of Aczel ([1])<sup>5</sup>

Fix an extensional combinatory algebra  $\mathcal{M}$ ; let  $|\mathcal{M}|$  be the universe of  $\mathcal{M}$ . If  $X = \langle X_0, X_1 \rangle$ ,  $Y = \langle Y_0, Y_1 \rangle$ , and  $X_0, X_1, Y_0, Y_1$  are subsets of  $\mathcal{M}$ , define

$$X \le Y \Leftrightarrow X_0 \subseteq Y_0 \land (\forall a \in X_0) (a \in X_1 \leftrightarrow a \in Y_1)$$

Let  $\mathcal{F}$  be the family of all pairs  $X = \langle X_0, X_1 \rangle$  of subsets of |M|, satisfying the restriction  $X_1 \subseteq X_0$ . If  $X \in \mathcal{F}$ , we call X suitable.

**Lemma 4.** The structure  $\langle \mathcal{F}, \leq \rangle$  is a complete partial ordering <sup>6</sup>.

We then define an operator  $\Gamma$  on suitable subsets of |M|.  $\Gamma(X)$  can be given by specifying two operators  $\Gamma_0(X)$ ,  $\Gamma_1(X)$  such that  $\Gamma(X) = \langle \Gamma_0(X), \Gamma_1(X) \rangle$ .  $\Gamma_0(X)$  is defined by the following formula  $A_0(x, X)$ :

$$\exists u \exists v \quad [ \quad (x = [u = v]) \lor (x = [Pu] \land u \in X_0) \lor \\ \lor \quad (x = [Tu] \land u \in X_0) \lor \\ \lor \quad (x = (\dot{\neg}u) \land x \in X_0) \lor \\ \lor \quad ((x = (u \dot{\lor}v) \lor x = (u \dot{\land}v)) \land u \in X_0 \land v \in X_0) \lor \\ \lor \quad (x = (u \dot{\rightarrow}v) \land u \in X_0 \land (u \notin X_1 \lor v \in X_0)) \lor \\ \lor \quad ((x = \dot{\forall}u \lor x = \dot{\exists}u) \land \forall x(ux) \in X_0)]$$

 $<sup>{}^{5}</sup>$ As an alternative, we can define the model by transfinite recursion over the ordinals, in much the same way as the model for Feferman's theories with the so-called join axiom.

<sup>&</sup>lt;sup>6</sup>I.e. a partial ordering in which every  $\leq$ -increasing sequence of elements of  $\mathcal{F}$  has a least upper bound with respect to  $\leq$ .

 $\Gamma_1(X)$  is defined by the following formula  $A_1(x, X)$ :

$$\exists u \exists v \quad [ \quad (x = [u = v] \land u = v) \lor \\ \lor \quad (x = [Pu] \land u \in X_0) \lor (x = [Tu] \land u \in X_1) \lor \\ \lor \quad (x = (\neg u) \land u \in X_0 \land u \notin X_1) \lor \\ \lor \quad (x = (u \dot{\lor} v) \land u \in X_0 \land v \in X_0 \land (u \in X_1 \lor v \in X_1)) \lor \\ \lor \quad (x = (u \dot{\land} v) \land u \in X_0 \land v \in X_0 \land u \in X_1 \land v \in X_1) \lor \\ \lor \quad (x = (u \dot{\rightarrow} v) \land u \in X_0 \land v \in X_0) \land (u \notin X_1 \lor v \in X_1)) \lor \\ \lor \quad (x = \dot{\lor} u \land \forall x(ux) \in X_0 \land \forall x(ux) \in X_1) \lor \\ \lor \quad (x = \dot{\exists} u \land \forall x(ux) \in X_0 \land \exists x(ux) \in X_1)]$$

The following properties are immediate:

**Lemma 5.** (i)  $\Gamma : \mathcal{F} \to \mathcal{F}$ . (ii)  $X \leq Y$ , then  $\Gamma(X) \leq \Gamma(Y)$  (where  $X, Y \in \mathcal{F}$ ). Hence there are sets  $X \in \mathcal{F}$ , such that  $X = \Gamma(X)$ .

**Theorem 3.** If  $X \in \mathcal{F}$  and  $X = \Gamma(X)$ , then

$$\langle \mathcal{M}, X \rangle \models \mathrm{PT}.$$

Hence PT is consistent.

The proof of the theorem is straightforward.

## 2.2. AT-models

As to the system AT, we first inductively define the set of propositional objects over a given combinatory algebra. We then exploit the stages assigned to propositional objects for generating the truth set.

Formally, we fix an extensional combinatory algebra  $\mathcal{M}$  with universe  $|\mathcal{M}|$ . We also assume that our language includes names for objects of  $\mathcal{M}$  (for which we adopt the same symbols). If t is a term of the expanded language,  $t^{\mathcal{M}}$ stands for the value of t in  $\mathcal{M}$ . We are now ready to define by transfinite recursion on ordinals a sequence  $\{\mathcal{P}_{\alpha}\}$  of subsets of  $|\mathcal{M}|$ :

• Initial clause:

$$\mathcal{P}_0 = \{ [a = b] \mid a, b \in M \} \cup \{ [Pa] \mid a \in M \};\$$

• Limit clause: if  $\lambda$  is a limit ordinal,

$$\mathcal{P}_{\lambda} = \cup \{ \mathcal{P}_{\alpha} \, | \, \alpha < \lambda \};$$

•  $\neg$ - and  $\wedge$ -rules:

$$\frac{a \in \mathcal{P}_{\alpha}}{(\dot{\neg} a)^M \in \mathcal{P}_{\alpha+1}} \qquad \frac{a \in \mathcal{P}_{\alpha} \quad b \in \mathcal{P}_{\alpha}}{(a \dot{\land} b)^M \in \mathcal{P}_{\alpha+1}}$$

•  $\forall$ - and *T*-rules:

$$\frac{\text{for all } c \in |M|, (fc)^M \in \mathcal{P}_{\alpha}}{\dot{\forall} f^M \in \mathcal{P}_{\alpha+1}} \qquad \frac{a \in \mathcal{P}_{\alpha}}{(\dot{T}a)^M \in \mathcal{P}_{\alpha+1}}$$

In a similar way, we recursively produce a sequence  $\{\mathcal{T}_{\alpha}\}$  of subsets of |M|, approximating the truth set :

• Initial clause:

$$\frac{\mathcal{M}\models a=b}{[a=b]^M\in\mathcal{T}_0};$$

• Limit clause: if  $\lambda$  is a limit ordinal,

$$\mathcal{T}_{\lambda} = \cup \{ \mathcal{T}_{\alpha} \mid \alpha < \lambda \};$$

• First successor clause: if  $a \in \mathcal{P}_{\alpha}$ ,

$$a \in \mathcal{T}_{\alpha+1} \Leftrightarrow a \in \mathcal{T}_{\alpha}$$

- Second successor clause: assume  $a \in \mathcal{P}_{\alpha+1} \mathcal{P}_{\alpha}$ . We distinguish several cases according to the form of a, i.e.  $a = \forall f, Tb, \neg b, b \land c$  (respectively).
  - 1.  $\forall$  and *T*-clauses:

$$\frac{\text{for all } c \in |M|, (fc)^M \in \mathcal{T}_{\alpha}}{\dot{\forall} f^M \in \mathcal{T}_{\alpha+1}} \qquad \frac{b \in \mathcal{T}_{\alpha}}{(\dot{T}b)^M \in \mathcal{T}_{\alpha+1}}$$

2.  $\wedge$ - and  $\neg$ -clauses:

$$\frac{b \in \mathcal{T}_{\alpha} \quad c \in \mathcal{T}_{\alpha}}{(b \dot{\wedge} c)^M \in \mathcal{T}_{\alpha+1}} \qquad \frac{b \notin \mathcal{T}_{\alpha}}{(\dot{\neg} b)^M \in \mathcal{T}_{\alpha+1}}$$

By transfinite induction on ordinals, it is not difficult to verify:

#### **Lemma 6.** If $a \in M$ , then

$$\begin{aligned} a \in \mathcal{T}_{\alpha} &\Rightarrow a \in \mathcal{P}_{\alpha}; \\ \alpha \leq \beta &\Rightarrow \mathcal{P}_{\alpha} \subseteq \mathcal{P}_{\beta} \wedge \mathcal{T}_{\alpha} \subseteq \mathcal{T}_{\beta}; \\ a \in \mathcal{P}_{\alpha} &\Rightarrow a \in \mathcal{T}_{\alpha} \lor (\dot{\neg} a) \in \mathcal{T}_{\alpha+1}; \\ (\dot{\neg} a) \in \mathcal{T}_{\alpha} &\Rightarrow a \notin \mathcal{T}_{\beta} (\beta \text{ arbitrary }). \end{aligned}$$

We choose  $\mathcal{P} = \bigcup \{\mathcal{P}_{\alpha} \mid \alpha \text{ ordinal }\}, \mathcal{T} = \bigcup \{\mathcal{T}_{\alpha} \mid \alpha \text{ ordinal }\}; \text{ of course, } \mathcal{P}, \mathcal{T} \text{ depend on the underlying combinatory algebra } \mathcal{M}, \text{ but we leave this fact implicit.}$ 

**Lemma 7.** Let  $\mathcal{O}$  be the open term model of combinatory logic plus extensionality <sup>7</sup>. Then it holds over  $\mathcal{O}$ :

$$\mathcal{P} = \mathcal{P}_{\omega} \text{ and } \mathcal{T} = \mathcal{T}_{\omega}$$

*Proof.* Assume that we have proved

$$\mathcal{P} = \mathcal{P}_{\omega}.\tag{1}$$

By lemma 6, if  $a \in \mathcal{T}_{\delta}$ , then  $a \in \mathcal{P}_{\delta}$ . By assumption,  $a \in \mathcal{P}_k$ , for some finite k. By the third claim of lemma 6, either  $a \in \mathcal{T}_k$  or  $(\neg a) \in \mathcal{T}_{k+1}$ . In the first case we are done; the second case implies  $a \notin \mathcal{T}_{\delta}$  (again lemma 6, last claim), contradiction. Hence  $\mathcal{T}_{\delta} \subseteq \mathcal{T}_{\omega}$ . But the converse inclusion trivially holds and hence  $\mathcal{T} = \mathcal{T}_{\omega}$ .

It remains to check (1). It is sufficient to define a recursively enumerable derivability relation  $\vdash$  over the term model, such that, for every  $a \in \mathcal{O}$ ,

 $\vdash a \Leftrightarrow a \in \mathcal{P}$ 

But this is straightforward: the axioms of  $\vdash$  will have the form [t = s], [P(t)], while the inference rules correspond to the positive inductive clauses generating the sequence  $\{\mathcal{P}_{\alpha}\}$ . Of course, the clause for  $\forall$  can be rephrased as a finitary inference: from  $\vdash ax$ , infer  $\vdash \forall a$ , provided x is not free in a.

It is then easy to check that the derivability relation is closed under substitution, that is, for arbitrary terms a, s:

$$\vdash a(x) \Rightarrow \vdash a[x := s]$$

This property together with the fact that  $\mathcal{O}$  is the open term model readily yields the initial claim (1)(proofs are carried out by induction on the definition of  $\vdash$  and by transfinite induction on ordinals). QED

#### **Theorem 4.** $\langle \mathcal{M}, \mathcal{P}, \mathcal{T} \rangle \models AT.$

**2.2.1. Proof-theoretic digression** Of course, we can consider applied versions of PT and AT. Indeed, let PTN (ATN) be PT (AT) extended with a predicate N for the set of natural numbers, constants for 0, successor, predecessor and conditional on N, the induction schema for natural numbers for N. Then:

**Theorem 5.** (i) PTN is proof-theoretically equivalent to ramified analysis of arbitrary level below  $\epsilon_0$ .

(ii) If N-induction is restricted to propositional functions, the resulting system  $PTN_c$  is proof-theoretically equivalent to Peano arithmetic.

<sup>&</sup>lt;sup>7</sup>For details see [2].

The proof follows well-known paths; the lower bound can be obtained by embedding in PTN a system of the required strength, for instance Feferman's  $EM_0+J$  [7].

As to the upper bound, it is possible to provide a proof-theoretic analysis of PTN with predicative methods (partial cut elimination and asymmetric interpretation into a ramified system with levels  $< \varepsilon_0$ ). A quick proof of the conservation result exploits recursively saturated models.

Concerning the strength of ATN, we do not have a definite result yet, but we believe that the following is true:

Conjecture 6. (i) ATN has the same proof theoretic strength of ACA, the system of second order arithmetic based on arithmetical comprehension.(ii) ATN with number theoretic induction restricted to propositional functions

is proof-theoretically equivalent to Peano arithmetics.

As to possible routes for proving (i)-(ii), one ought to consider lemma 7 and the methods of Glass [10].

### 3. Stratified Truth ?

We now explore an alternative route, which takes into account the possibility of dealing with the paradox in a fully impredicative, extensional framework, Quine's set theory NF. In the new model, the set of all propositions and the set of all truths do exist, and, to a certain extent, the notion of truth has rather strong closure properties.

We first describe the formal details.  $\mathcal{L}_s$  is the elementary set theoretic language, which comprises the binary predicate symbol  $\in$ .  $\mathcal{L}_s$ -terms are simply individual variables (x, y, z, ...) and prime formulas (atoms) have the form  $t \in$ s (t, s terms).  $\mathcal{L}_s$ -formulas are inductively generated from prime formulas by means of sentential connectives and quantifiers. The elementary set theoretic language  $\mathcal{L}_s^+$  is obtained by adding to  $\mathcal{L}_s$  the abstraction operator  $\{-|-\}$ ;  $\mathcal{L}_s^+$ -terms and formulas are then simultaneously generated. The clause for introducing class terms has the form: if  $\varphi$  is a formula, then  $\{x | \varphi\}$  is a term where  $FV(\{x | \varphi\}) = FV(\varphi - \{x\})$  (FV(E) is the set of free variables occurring in the expression E).

As usual, two terms (formulas) are called  $\alpha$ -congruent if they only differ by renaming of bound variables; we identify  $\alpha$ -congruent terms (formulas).

#### 3.1. Stratified comprehension

As usual for Quine's systems, we need the technical device of *stratification*; we also define a restricted notion thereof, which is motivated by the consideration of "loosely predicative" class existence axioms.

- (i)  $\varphi$  is *stratified* iff it is possible to assign a natural number (type in short) to each term occurrence <sup>8</sup> of  $\varphi$  in such a way that
  - 1. if  $t \in s$  is a subformula of  $\varphi$ , the type of s is one greater than the type of t;
  - 2. all free occurrences of the same variable in any subformula of  $\varphi$  have the same type;
  - 3. if x is free in  $\psi$  and  $\forall x\psi$  is a subformula of  $\varphi$ , then the 'x' in  $\forall x$  and the free occurrences of x in  $\psi$  receive the same type;
  - 4. if  $t = \{x \mid \beta\}$  occurs in  $\varphi$ , x is free in  $\beta$ , then t is assigned a type one greater than the type assigned to x, and all the free occurrences of x in  $\beta$  receive the same type.
- (ii)  $\{x \mid \varphi\}$  is stratified if  $\varphi$  is stratified;
- (iii) a stratified term  $\{x \mid \varphi(x, \vec{y})\}$  is *loosely predicative* iff for some type  $i \in \omega$ ,  $\{x \mid \varphi(x, \vec{y})\}$  has type i + 1, no (free or bound) variable of  $\varphi(x, \vec{y})$  is assigned type greater than i+1; a stratified term  $\{x \mid \varphi(x, \vec{y})\}$  is *predicative* iff  $\{x \mid \varphi(x, \vec{y})\}$  is loosely predicative and in addition no quantified variable of  $\varphi(x, \vec{y})$  is assigned the same type as  $\{x \mid \varphi(x, \vec{y})\}$  itself.

(iv)  $\varphi$  is n + 1-stratified iff  $\varphi$  is stratified by means of  $0, \ldots, n$ .

For instance,  $\bigcup a = \{x \mid (\exists y \in a) (x \in y)\}$  is not loosely predicative, since it requires type 2, but  $\bigcup a$  itself has type 1;  $a \cap b = \{x \mid x \in a \land x \in b\}$  is predicative.

**Definition 4.** The system NF comprises:

1. predicate logic for the extended language  $^9$ ;

$$\forall u(\varphi(u) \leftrightarrow \psi(u)) \rightarrow \{x \,|\, \varphi(x)\} = \{x \,|\, \psi(x)\}.$$

This would ensure that  $\{x \mid \neg x = x\} = \{x \mid x \in x \land \neg x \in x\}$ . An alternative route would be to extend the logic with a description operator, say in the style of [14], ch.VII. If this choice is adopted, the previous schema becomes provable. Be as it may, the resulting theories are conservative over NF as formalized in the pure set theoretic language  $\mathcal{L}_s$ , and we won't bother the reader with further details.

 $<sup>^{8}\</sup>mathrm{Individual}$  constants included; these can be given any type compatible with the clauses below.

 $<sup>^{9}</sup>$  To be more accurate, if the abstraction operator is assumed as primitive, it is convenient to include in the extended logic the schema

2. class extensionality:  $\forall x \forall y (x =_e y \rightarrow x = y)$ , where

$$\begin{split} t &= s :\Leftrightarrow \forall z (t \in z \to s \in z); \\ t &=_e s :\Leftrightarrow \forall x (x \in t \leftrightarrow x \in s) \end{split}$$

3. stratified explicit comprehension SCA: if  $\varphi$  is stratified, then

$$\forall u(u \in \{x \,|\, \varphi(x, \vec{y})\} \leftrightarrow \varphi(u, \vec{y}))$$

#### Other systems

- (i) NFP (NFI) is the subsystem of NF, where SCA is restricted to (loosely) predicative abstracts.
- (ii)  $NF_k$  (NFI<sub>k</sub>, NFP<sub>k</sub>) is the subsystem of NF (NFI, NFP<sub>k</sub>), where (at most) k types are allowed for stratification.

**Remark 2.** If Union is the axiom " $\bigcup a$  exists, for all a ", then NFP+Union is equivalent to the full NF (cf. [6])

#### 3.2. Consistent NF-subtheories

By a theorem of Crabbè ([6]), NFI is provably consistent in third order arithmetic. The details of the (different) consistency proofs for NFI can be found in [6] and [13]. The main idea of [6], pp.133-135, is to exploit the reducibility of NFI to its fragment NFI<sub>4</sub>; then NFI<sub>4</sub> is interpreted into a corresponding type theory up to level 4, TI<sub>4</sub>, plus Amb (= the so called schema of typical ambiguity, see [20])<sup>10</sup> These steps are finitary and adapt well-known theorems of Grishin and Specker. TI<sub>4</sub>+Amb is shown consistent by means of the Hauptsatz for second order logic (which is equivalent in primitive recursive arithmetic to the 1-consistency of full second order arithmetic<sup>11</sup>). Hence:

**Theorem 7.** NFI is consistent (in primitive recursive arithmetic plus the 1– consistency of second order arithmetic).

In order to carry out a Kripke-like construction in the NF-systems and to represent the syntax, we shall essentially exploit Quine's homogeneous pairing operation, which *does require extensionality* and the existence of a copy of the natural numbers. But it is not difficult to check that Quine's pairing is indeed well-defined already in NFI.

<sup>&</sup>lt;sup>10</sup>In TI<sub>4</sub>, { $x^i | \varphi$ } exists, provided i = 0, 1, 2 and  $\varphi$  contains free or bound variables of type i + 1 at most. <sup>11</sup>E.g. see J.Y. Girard, *Proof Theory and Logical Complexity*, Bibliopolis, Napoli 1987,

<sup>&</sup>lt;sup>11</sup>E.g. see J.Y. Girard, *Proof Theory and Logical Complexity*, Bibliopolis, Napoli 1987, p.280.

This requires two steps. First of all, the collection of Fregean natural numbers is a set in NFI. Define:

$$\begin{split} \emptyset &= \{x \mid x \neq x\}; \\ V &= \{x \mid x = x\} \\ 0 &= \{\emptyset\}; \\ a+1 &= \{x \cup \{y\} \mid x \in a \land y \notin x\}; \\ Cl_N(y) &\Leftrightarrow 0 \in y \land \forall x (x \in y \to (x+1) \in y); \\ \mathcal{N} &= \{x \mid \forall y (Cl_N(y) \to x \in y)\} \end{split}$$

Then NFI grants the existence of  $\mathcal{N}$ ; in fact, by inspection, all the above sets above are loosely predicative. Furthermore, we have, provably in NFI:

#### Lemma 8 (NFI).

$$Cl_N(\{x \mid \varphi(x)\}) \to \mathcal{N} \subseteq \{x \mid \varphi(x)\};$$
<sup>(2)</sup>

$$Cl_{N}(\{x \mid \varphi(x)\}) \to \mathcal{N} \subseteq \{x \mid \varphi(x)\};$$

$$(\forall x)(x \in \mathcal{N} \leftrightarrow x = 0 \lor (\exists y \in \mathcal{N})(x = y + 1));$$

$$(\exists x \in \mathcal{N} \land (\forall x \in \mathcal{N})(V \notin x);$$

$$(A)$$

$$\emptyset \notin \mathcal{N} \land (\forall x \in \mathcal{N})(V \notin x); \tag{4}$$

$$(\forall x \in \mathcal{N})(x+1 \neq 0); \tag{5}$$

$$(\forall x \in \mathcal{N})(\forall y \in \mathcal{N})(x+1=y+1 \to x=y) \tag{6}$$

(In (2)  $\{x \mid \varphi(x)\}$  must be loosely predicative).

Clearly  $\mathcal{N}$  is infinite by (4) above. As to the proof, (4) holds in NFI + Union, as  $NFI+Union \equiv NF$ , and NF proves (4) according to a famous result of Specker ([19]). On the other hand,  $NFI + \neg Union$  implies (4) by [6]. The claims (3), (2) with the Peano axioms are provable in NFI ((6) requires the second part of (4)).

**Definition 5.** (Homogeneous pairing; [18])

$$\begin{array}{lll} \phi(a) &=& \{y \,|\, y \in a \land y \notin \mathcal{N}\} \cup \{y+1 | y \in a \land y \in \mathcal{N}\};\\ \theta_1(a) &=& \{\phi(x) \,|\, x \in a\};\\ \theta_2(a) &=& \{\phi(x) \cup \{0\} \,|\, x \in a\};\\ (a,b) &=& \theta_1(a) \cup \theta_2(b);\\ Q_1(a) &=& \{z \,|\, \phi(z) \in a\};\\ Q_2(a) &=& \{z \,|\, \phi(z) \cup \{0\} \in a\}\end{array}$$

The definitions above are (at most) loosely predicative and hence the universe of sets is closed under the corresponding operations, provably in NFI.

**Lemma 9** (NFI). 1.  $\phi(a) = \phi(b) \rightarrow a = b$ ;

- 2.  $0 \notin \phi(a);$
- 3.  $\theta_i(a) = \theta_i(b) \rightarrow a = b$ , where i = 1, 2;
- 4.  $Q_i((x_1, x_2)) = x_i$ , where i = 1, 2.
- 5.  $(x, y) = (u, v) \rightarrow x = u \land y = v$ .
- 6. the map  $x, y \mapsto (x, y)$  is surjective and  $\subseteq$ -monotone in each variable.

The proof hinges upon the properties of  $\mathcal{N}$  and the successor operation ([18]). In particular, we below exploit the fact that Quine's pairing operation is  $\subseteq$ -monotone in both arguments. This is seen by inspection: the definition of (a, b) is positive in  $a, b^{12}$ .

**Lemma 10** (Fixed point). Let A(x, a) be a formula which is positive in a. Assume that

$$\Gamma_A(a) = \{x \,|\, A(x,a)\}$$

is loosely predicative, where x, a are given types i, i+1 respectively. Then NFI proves the existence of a set c of type i + 1, such that:

- $\Gamma_A(c) \subseteq c$ ;
- $\Gamma_A(a) \subseteq a \Rightarrow c \subseteq a$ .

The proof is standard: observe that the set

$$c := \{ x \, | \, \forall d(\Gamma_A(d) \subseteq d \to x \in d) \}$$

is loosely predicative  $^{13}$ .

**Definition 6.** NFI(pair) (NFP(pair), NF(pair)) is the theory, which extends NFI (NFP, NF respectively) with a new binary function symbol (-, -) for ordered pairing and the corresponding new axiom:

$$\forall x \forall y \forall u \forall v ((x, y) = (u, v) \to x = u \land y = v)$$

It is understood that the stratification condition is lifted to the new language by stipulating that t, s in (t, s) receive the same type.

The strategic role of homogeneous pairing is clarified by two equivalence results. Below let  $S_1 \equiv S_2$  denote the relation that holds between two formal theories  $S_1$ ,  $S_2$  whenever they are mutually interpretable.

<sup>&</sup>lt;sup>12</sup>We recall that a formula A(x, a) is positive in a if every free occurrence of a in the negation normal form of A is located in atoms of the form  $t \in a$ , which are prefixed by an even number of negations and where  $a \notin FV(t)$ .

 $<sup>^{13}\</sup>mathrm{A}$  similar argument shows that NFI justifies the existence of the largest fixed point of  $\Gamma_A.$ 

#### **Proposition 6.** $NF \equiv NF_3(pair)$

The proof of this statement was independently suggested by Antonelli and Holmes. The basic observation is that NFP<sub>3</sub>(pair) proves the existence of the set E, where  $\mathcal{E} := \exists y(y = E)$  and  $E = \{\{\{x\}, y\} | x \in y\}$ . But by Grišin [12], NF  $\equiv$  NF<sub>3</sub> +  $\mathcal{E}$ .

We can readily extend the proposition to the subsystems NFI, NFP.

**Proposition 7.** NFP  $\equiv$  NFP<sub>3</sub>(pair) and NFI  $\equiv$  NFI<sub>3</sub>(pair)

Proof. Let

$$\mathcal{I} \Leftrightarrow \exists y (y = \{u \mid \cap u \neq \emptyset\})$$

By a theorem of Grishin and [6],  $\mathsf{NFI} \equiv \mathsf{NFI}_3 + \mathcal{I}$ , (respectively  $\mathsf{NFP} \equiv \mathsf{NFP}_3 + \mathcal{I}$ ). But  $\mathsf{NFP}_3(\mathsf{pair})$  proves that  $M = \{(\{x\}, y) | x \in y\}$  is a set and

$$\forall s \exists a \forall x (x \in a \leftrightarrow \exists q (q \in s \land (\forall z \in x) ((q, z) \in M))) \tag{7}$$

Choose  $s = \{\{x\} | x \in V\}$  in (7); then NFP<sub>3</sub>(pair) proves that there is a set I such that

$$(\forall x)(x \in I \leftrightarrow \cap x \neq \emptyset)$$

It follows that we can freely use the homogeneous pairing operation when we work in  $\mathsf{NF}$  and  $\mathsf{NFI}.$ 

# 4. Generating a truth set in NFI

We now simulate via homogeneous pairing the logical operations, which are essential for introducing in the Quinean universe a counterpart of the formula-representing map of  $\S1.1$ .

#### Definition 7.

$$\begin{array}{rcl} \dot{\neg}x & := & (0,x); \\ x \dot{\wedge}y & := & (1,(x,y)); \\ \dot{\forall}f & := & (2,f); \\ \dot{\in}xy & := & (3,(x,y)) \end{array}$$

Then we can inductively introduce a map  $A \mapsto [A]$  such that FV(A) = FV([A]) and

- $[t \in s] := \dot{\in} \{t\}s;$
- $[\neg A] := \dot{\neg}[A]$ ;
- $[A \wedge B] := \dot{\wedge}[A][B]$ ;
- $[\forall xA] := \dot{\forall} \{x \mid A\}.$

We also define

$$y \cdot x := [x \in y]$$

Under the dot-application, the universe of sets becomes an applicative structure. Note however that  $y \cdot x$  is stratified only if y and x are given the types i + 1 and i (respectively), and that the result of applying y to x is one greater than the type of x.

Lemma 11 (Restricted diagonalization, in NFI). There is a term R such that

 $R=\dot{\neg}R$ 

Moreover, for every a, there is a term  $\Delta(a)$  such that

$$\Delta(a) = [\neg(a \in \Delta(a))]$$

*Proof.* The operation  $\Gamma(a) = (\neg, a)$  is monotone in a. By the fixed point lemma, there is some R satisfying the claim. As to the second part,  $\Phi_a(b) = [a \in b] = (3, (a, b))$  is monotone in b. So the conclusion is implied by lemma 10.

We now model the Kripke-Feferman notion of self-referential truth, as developed in [3], within the abstract framework of Quine's set theory. The truth predicate  $\mathcal{T}$  is introduced as the fixed point of a stratified positive (in *a*) operator  $\mathcal{T}(x, a)$ , which encodes the recursive clauses for partial self-referential truth and is given by the formula

$$\exists u \exists v \exists w \quad [ \quad (x = [u \in v] \land u \in v) \lor \lor (x = [\neg u \in v] \land \neg u \in v) \lor \lor \lor (x = [\neg \neg v] \land v \in a) \lor \lor (x = v \land w \land v \in a \land w \in a) \lor \lor (x = \dot{\neg} (v \land w) \land (\dot{\neg} v \in a \lor \dot{\neg} w \in a)) \lor \lor (x = \dot{\neg} v \land \forall z (v \cdot z \in a)) \lor \lor (x = \dot{\neg} \forall v \land \exists z (\dot{\neg} v \cdot z \in a)))$$

Clearly  $\{x \mid \mathcal{T}(x, a)\}$  is  $\subseteq$ -monotone in a and is predicative: it receives type 2 once we assign type 0 to u, z, type 1 to x, v, w, type 2 to a.

Definition 8.

$$Cl_T(a) := \forall x(\mathcal{T}(x, a) \to x \in a);$$
  

$$T := \{x \mid \forall a(Cl_T(a) \to x \in a)\};$$
  

$$Ta : := a \in T.$$

Lemma 10 immediately implies:

Proposition 8. NFI proves:

- 1.  $\exists y(y=T);$
- 2.  $\forall a(\mathcal{T}(a,T) \to a \in T);$
- 3.  $Cl_T(a) \to T \subseteq a$ .

Proposition 9. NFI proves:

$$\begin{array}{rcccc} T[x \in y] & \leftrightarrow & x \in y; \\ T[\neg x \in y] & \leftrightarrow & \neg x \in y; \\ T \dot{\neg} \dot{\neg} \dot{\neg} x & \leftrightarrow & Tx; \\ T(x \dot{\wedge} y) & \leftrightarrow & Tx \wedge Ty; \\ T \dot{\neg} (x \dot{\wedge} y) & \leftrightarrow & T(\dot{\neg} x) \vee T(\dot{\neg} y); \\ T \dot{\forall} f & \leftrightarrow & \forall x T(f \cdot x); \\ T \dot{\neg} \dot{\forall} f & \leftrightarrow & \exists x T \dot{\neg} (f \cdot x). \end{array}$$

*Proof.* Use  $Ta \leftrightarrow \mathcal{T}(a, T)$ , which is provable with proposition 8, and the independence properties of  $\dot{\neg}, \dot{\land}, \dot{\forall}, \dot{\in}$ , which follow from lemmata 8–9. QED

Hence, as interesting special cases, we obtain:

$$(T[Tx] \leftrightarrow Tx) \land (T[\neg Tx] \leftrightarrow \neg Tx)$$

Proposition 10 (Consistency). NFI proves

- 1.  $\forall x(T \dot{\neg} x \rightarrow \neg T \dot{\neg} x);$
- 2.  $\exists y(\neg Ty \land \neg T \dot{\neg} y).$

*Proof.* Ad 1: choose  $\psi(a) := \neg T(\neg a)$ ;  $\{x \mid \psi(x)\}$  exists in NFI. Then check

$$\forall a(\mathcal{T}(a, \{x \,|\, \psi(x)\}) \to \psi(a))$$

Ad 2: let y = R (lemma 11) and apply consistency.

Consider the map

$$x \mapsto \varphi(x) := [\neg Tx]$$

**Remark 3.** An alternative "Liar propositional object" would be a set L such that

$$L = [\neg TL] = [\neg (L \in T)].$$

But observe that the above equation cannot be stratified. Indeed, NFI proves

$$\neg \exists x (x = [\neg Tx])$$

**Lemma 12.** If A is stratified (and  $\{x \mid A\}$  is loosely predicative), then

$$T[\forall xA] \quad \leftrightarrow \quad \forall xA;$$
  
$$T[\neg \forall xA] \quad \leftrightarrow \quad \neg \forall xA$$

are provable in NF (NFI).

*Proof.* Consider the following steps:

$$T[\forall xA] \quad \leftrightarrow \quad \forall uT(\{x \mid A\} \cdot u);$$
  
$$\leftrightarrow \quad \forall uT[u \in \{x \mid A\}];$$
  
$$\leftrightarrow \quad \forall u(u \in \{x \mid A\});$$
  
$$\leftrightarrow \quad \forall uA[x := u]$$

Observe that the second step uses proposition 9, while the last step requires stratified comprehension in NF (or in NFI, provided A is loosely predicative).  $\Box$ 

Theorem 8 (Stratified T-schema). If A is stratified, NF proves

 $T[A] \leftrightarrow A$ 

If A is  $\forall$ -free, the schema is already provable in NFI.

*Proof.* By induction on A with proposition 9 and the previous lemma. QED

The stratified T-schema implies that T strongly deviates from the behaviour of self-referential truth predicates à la Kripke-Feferman, which cannot in general apply to the truth axioms themselves nor to *arbitrary* logical axioms (see [15], [8]). On the contrary, T provably believes that it is two-valued, consistent and that it satisfies the closure conditions embodied by the operator formula T(x,T) generating partial truth itself:

# Corollary 2 (vs. KF). NFI proves:

$$\begin{array}{rcl} T[Ta & \lor & \neg Ta]; \\ T[\neg(Ta & \land & \neg Ta)]; \\ T[T[x \in y] & \leftrightarrow & x \in y]; \\ T[T[\neg x \in y] & \leftrightarrow & \neg x \in y]; \\ T[T \dot{\neg} \dot{\neg} x & \leftrightarrow & Tx]; \\ T[T(x \dot{\land} y) & \leftrightarrow & Tx \land Ty]; \\ T[T \dot{\neg} (x \dot{\land} y) & \leftrightarrow & T(\dot{\neg} x) \lor T(\dot{\neg} y)] \end{array}$$

In addition,  $\mathsf{NF} \vdash T[\forall xT(f \cdot x) \leftrightarrow T \dot{\forall} f] \land T[\forall x(Tx \leftrightarrow \mathcal{T}(x,T))].$ 

Proof. As to the first part, apply proposition 9 and theorem 7. Concerning the second part, first observe:

$$\begin{array}{lll} \forall x(f \cdot x \in T) & \equiv & \forall xT(f \cdot x) \\ & \leftrightarrow & \forall x(x \in \{u \mid f \cdot u \in T\}); \\ & \leftrightarrow & \forall xT[x \in \{u \mid f \cdot u \in T\}]; \\ & \leftrightarrow & \forall xT(\{u \mid f \cdot u \in T\} \cdot x); \\ & \leftrightarrow & T \forall \{u \mid f \cdot u \in T\}; \\ & \leftrightarrow & T [\forall u T(f \cdot u)] \end{array}$$

Hence

$$\begin{aligned} \neg T(\dot{\forall}f) \lor \forall x T(f \cdot x) &\Rightarrow T[\neg T(\dot{\forall}f)] \lor T[\forall x T(f \cdot x)] \\ &\Rightarrow T[\neg T(\dot{\forall}f) \lor \forall x T(f \cdot x)] \\ &\Rightarrow T[T(\dot{\forall}f) \to \forall x T(f \cdot x)] \end{aligned}$$

Similarly, using

$$\neg \forall x T(f \cdot x) \leftrightarrow T[\neg \forall x T(f \cdot x)]$$

we obtain  $T[\forall x T(f \cdot x) \to T \dot{\forall} f]$ .

# 4.1. Final remarks

# Definition 9.

$$Pa:\Leftrightarrow Ta\vee T\dot{\neg}a$$

Pa formally represents the predicate "a is a proposition".

**Proposition 11** (NFI). The collection of all propositions is a proper subset of the universe:

$$\exists y(y = \{x \mid Px\}) \land \{x \mid Px\} \subset V$$

Moreover P has the following closure properties:

$$P([x \in y]) \land P[Tx];$$

$$Pa \land (Ta \to Pb) \to P(a \dot{\rightarrow} b);$$

$$Pa \land Pb \to P(a \dot{\land} b) \land P(a \dot{\lor} b);$$

$$Pa \to T[Pa];$$

$$P(\dot{\forall} f).$$

The first claim is a consequence of proposition 10. As to the remaining properties, apply proposition 9.

A few closure conditions of the previous proposition are reminiscent of corresponding axioms in PT and AT; but we stress that the axiom  $T[\neg Px]$  of AT is refuted in the present context, while the (analogue of the) last statement is clearly unsound, provably in AT and PT. Note also that  $P(\forall f)$  implies  $\forall xP(f \cdot x)$ , i.e. every set defines a propositional function.

We conclude by representing Russell's contradiction within the theory of propositions and truth that we have sofar developed in NFI.

#### Definition 10.

$$\tau(f) := [P(\dot{\forall}f)]$$

By definition of the map  $A \mapsto [A]$ , pairing, and proposition 11, we obtain:

Lemma 13. NFI proves :

$$P(\tau(f)) \wedge T(\tau(f));$$
  
$$\tau(f) = \tau(g) \rightarrow f = g$$

So the operation  $\tau$  is a well-defined injective map from sets into truths (and propositions).

### Proposition 12. NFI proves:

$$\neg \exists d \forall x (x \in d \leftrightarrow (\exists f \subseteq P) (x \notin f \land x = \tau(f)))$$

To sum up: we have considered three formal systems for dealing with Russell's contradiction in Appendix B of [21]. In all systems the Russellian argument can be naturally formalized, but it is essentially sterilized either by denying the existence of a suitable propositional function or a set. The first two systems are provably non-extensional and predicatively inclined; no analogue of the power

set is apparently definable, while the collections of truths and propositions do not define completed totalities.

In the third system we do have the set of all propositions and truths and extensionality is basic. The contradiction is then avoided by the mechanism of stratification and in accordance with an impredicative type theoretic perspective.

#### References

- P. Aczel, Frege structures and the notions of proposition, truth and set, in: J.Barwise, H.J. Keisler, K. Kunen (eds.), *The Kleene Symposium*, North Holland, Amsterdam 1980,31-59.
- [2] H. Barendregt, The Lambda Calculus: its Syntax and Semantics, North Holland, Amsterdam 1984.
- [3] A. Cantini, *Logical Frameworks for Truth and Abstraction*, North Holland, Amsterdam 1996.
- [4] A.Cantini and P.Minari, Uniform inseparability in explicit mathematics, The Journal of Symbolic Logic, 64, 1999, 313-326.
- [5] N. Cocchiarella, Russell's paradox of the totality of propositions, Nordic Journal of Philosophical Logic, vol.5,2000, pp.23-37.
- [6] M. Crabbè, On the consistency of an impredicative subsystem of Quine's NF, The Journal of Symbolic Logic, 47, 1982, 131-136.
- S. Feferman, Constructive Theories of functions and classes, in: Logic Colloquium '78 (M.Boffa et alii, eds.), North Holland, Amsterdam, 1978, 55-98
- [8] S. Feferman, Reflecting on incompleteness, Journal of Symbolic Logic, 56, 1991, 1-49.
- [9] T. Forster, Set Theory with a Universal Set, Oxford Logic Guides, no.31, Oxford, 1995.
- [10] T. Glass, On power class in explicit mathematics, Journal of Symbolic Logic, 61, 1996, 468-489.
- [11] P. Grim, *The incomplete universe. Totality, Knowledge and Truth*, MIT Press, Cambridge Mass., 1991.
- [12] V.N.Grishin, Consistency of a fragment of Quine's NF system, Soviet Mathematics Doklady, vol.10 (1969), pp.1387-1390.
- [13] M.R.Holmes, The equivalence of NF-style set theories with "tangled" type theories: the construction of  $\omega$ -models of predicative NF (and more), *The Journal of Symbolic Logic*, vol.60 (1995), 178–190.
- [14] D. Kalish, R. Montague, Logic. Techniques of Formal Reasoning, Harcourt, Brace and World, Inc., New York, 1964.

- [15] S. Kripke, Outline of a theory of truth, *Journal of Philosophy*, 72, 1975, 690-716.
- [16] M. Oksanen, The Russell-Kaplan paradox and other modal paradoxes; a new solution, Nordic Journal of Philosophical Logic, vol.4, pp.73-93.
- [17] W.V.O. Quine , On ordered pairs, The Journal of Symbolic Logic, 10, 1945, 95-96.
- [18] J. B. Rosser, Logic for Mathematicians, Mc Graw-Hill, New York 1953.
- [19] E.Specker, The axiom of choice in Quine's New foundations for mathematical logic, *Proceedings of the National Academy of Sciences of the U.S.A.*, vol.39, 1953, 972-975.
- [20] E.Specker, Typical ambiguity, in: E.Nagel, P.Suppes and A.Tarski (eds.), Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress, Stanford University Press, Stanford 1962, 116-123.
- [21] B. Russell, The principles of Mathematics, London 1903 (reprinted by Routledge, London 1997).

Dipartimento di Filosofia Università degli Studi di Firenze via Bolognese 52, 50139 Firenze, Italy

Email: cantini@philos.unifi.it