A Simplification of Girard's Paradox

Antonius J.C. Hurkens

Klaasstokseweg 7, 5443 NS Haps, The Netherlands
e-mail: hurkens@sci.kun.nl

Abstract. In 1972 J.-Y. Girard showed that the Burali-Forti paradox can be formalised in the type system U. In 1991 Th. Coquand formalised another paradox in U^-. The corresponding proof terms (that have no normal form) are large. We present a shorter term of type $\bot$ in the Pure Type System λU^- and analyse its reduction behaviour. The idea is to construct a universe $\mathcal{U}$ and two functions such that a certain equality holds. Using this equality, we prove *and* disprove that a certain object in $\mathcal{U}$ is well-founded.

1 Introduction

Jean-Yves Girard (1972) derived a contradiction in the type system U by formalising a paradox inspired by those of Burali-Forti and Russell. By formalising another paradox, Thierry Coquand (1994) showed that the type system U^- is also inconsistent. So there are large proof terms of type $\bot$ in these type systems.

In Section 3 we present a relatively short term of type $\bot$ in λU^-. This Pure Type System and some notation is described in Section 2. In the last section we show that the β-reduction behaviour of the proof term is very simple.

In the other sections we will see that the proof has the same ingredients as Burali-Forti's paradox: a universe $\mathcal{U}$, a relation $<$ on $\mathcal{U}$, an object Ω in $\mathcal{U}$, and the question whether Ω is well-founded or not.

In Section 4 we describe Burali-Forti's paradox and some simplifications. We analyse the connection between the universe of all ordinals at its power set. In Section 5 we introduce *paradoxical* universes. These are connected to their power set in such a way that we can derive a Burali-Forti like contradiction. This can be formalised in Pure Type Systems. The formalisation can be simplified by considering *powerful* universes. In Section 6 we see how these universes are connected to the power set of their power set.

2 Pure Type Systems

In this section, we describe some Pure Type Systems. For more details, see for example (Barendregt 1992) or (Geuvers 1993).

2.1 The Pure Type Systems λHOL, λU^-, and λU

The typed λ-calculus λHOL (*Higher Order Logic*) is the Pure Type System (with β-conversion) given by the *sorts* $*$, $\Box$, and $\triangle$, the *axioms* $* : \Box$ and $\Box : \triangle$, and the *rules* $(*,*)$, $(\Box,\Box)$, and $(\Box,*)$. It is a consistent system, even if one adds the rule $(\triangle,*)$. Adding the rule $(\triangle,\Box)$, one gets the Pure Type System λU^-. Adding both rules, one gets λU.

2.2 Typing Terms in a Pure Type System

Each term A in a Pure Type System is either a variable x, a sort s, a product $\Pi x : B.\, C$, an abstraction $\lambda x : B.\, C$, or an application $(B\ C)$.

By $B[C/x]$ we denote the result of substituting the term C for the free occurrences of the variable x in B (renaming bound variables if necessary). By $=_\beta$ we denote the equivalence relation between terms that is induced by β-reduction: replacing a subterm of the form $(\lambda x : A.\, B\ C)$ by the term $B[C/x]$. If a term does not contain such a subterm, then the term is called *normal.*

In a Pure Type System, we can derive formal judgements $x_1 : A_1, \cdots, x_n : A_n \vdash B : C$, expressing that B *has type* C in the given context, that is, assuming that for $i = 1, \ldots, n$, variable x_i has type A_i.

We start in the empty context. If, in some context, A has type s for some sort s, then we are allowed to introduce a new variable x of type A.

The context gives the types of some variables.

The axioms give the types of some sorts.

We use the rules (s', s) to type products as follows: if A has type s' and (under the extra assumption $x : A$) B has type s, then (in the original context) $\Pi x : A.\, B$ also has type s.

If $\Pi x : A.\, B$ has type s and (under the extra assumption $x : A$) C has type B, then (in the original context) $\lambda x : A.\, C$ has type $\Pi x : A.\, B$.

If F has type $\Pi x : A.\, B$ and C has type A, then $(F\ C)$ has type $B[C/x]$.

Finally, we use β-reduction to change types: if A has type B, $B =_\beta C$, and C has type s, then we may conclude that A has type C.

Note that if a variable, abstraction or application has type A, then A is of type s for some sort.

2.3 Some Useful Properties of λU

Two terms A and B are β-equal if and only if for some C, both A and B reduce to C. If term B has a type D, then this type is unique up to β-equality. Furthermore, if B β-reduces to C, then C is also a term of type D.

We can calculate the *level* of a term (and its subterms) in a given context $x_1 : A_1, \cdots, x_n : A_n$ as follows: The sorts $*$, $\Box$, and $\triangle$ have level 2, 3, and 4, respectively. The level of variable x_i is one less then the level of A_i in the context $x_1 : A_1, \cdots, x_{i-1} : A_{i-1}$. The level of a product $\Pi x : B.\, C$ or an abstraction $\lambda x : B.\, C$ is the level of C in the extended context $x_1 : A_1, \cdots, x_n : A_n, x : B$. The level of an application $(B\ C)$ is the level of B in the original context.

One can prove that if B has type C in some context, then the level of B is one less than that of C. So each term has level 0, 1, 2, 3, or 4. One can also show that no term in λU contains a subterm of lower level (in the corresponding context). This implies that if we use a rule (s, s) to form a product $\Pi x : B.C$, then $\mathrm{level}(x) < \mathrm{level}(B) = \mathrm{level}(s) - 1 = \mathrm{level}(C)$, so the variable x has no free occurrence in C.

It turns out that each term A of level 1 is *strongly normalising*: there is no infinite sequence $A \to_\beta A' \to_\beta A'' \to_\beta \cdots$ of β-reduction steps. The terms of higher level are normal, since each abstraction or application has level 0 or 1.

2.4 The Five Levels of Terms in λU

We describe the five levels and introduce some notation to distinguish terms of different levels.

The only term of level 4 is Δ and the only term of level 3 is $\Box$.

We will call the terms of level 2 *sets* or *universes*. We think of $*$ as the *set* of all propositions. We use calligraphic letters $\mathcal{X}, \ldots$ for set variables.

We will call the terms of level 1 *objects*. Objects $\varphi, \chi, \ldots$ of type $*$ are called *propositions*. We use italic letters $x, \ldots$ for object variables.

Finally, the terms of level 0 are called *proofs* or *proof terms*. We use natural numbers $0, 1, \ldots$ for proof variables. These correspond exactly to the labels of assumptions in a natural deduction in Gentzen's style.

Using the rule $(\Box, \Box)$, we can form the *set* of all functions from a set $\mathcal{S}$ to a set $\mathcal{T}$:

$$(\mathcal{S} \to \mathcal{T}) \equiv \Pi x : \mathcal{S}.\, \mathcal{T}$$

In particular, the *power set* of $\mathcal{S}$ can be seen as the set of all predicates on $\mathcal{S}$:

$$\wp\mathcal{S} \equiv (\mathcal{S} \to *)$$

Using the rule $(\Delta, \Box)$, which is not allowed in λHOL, we can form a 'polymorphic domain' $\Pi\mathcal{X} : \Box.\, \mathcal{T}$ (where $\mathcal{X}$ may occur in $\mathcal{T}$). This product of level 2 has no clear set-theoretical interpretation. The products corresponding to the rules $(*, *)$, $(\Box, *)$, and $(\Delta, *)$ are propositions:

$$\begin{aligned}
[\varphi \Rightarrow \chi] &\equiv \Pi 0 : \varphi.\, \chi \\
\forall x : \mathcal{S}.\, \chi &\equiv \Pi x : \mathcal{S}.\, \chi \\
\forall \mathcal{X} : \Box.\, \chi &\equiv \Pi \mathcal{X} : \Box.\, \chi
\end{aligned}$$

Other connectives can be defined as usual. We only need falsehood and negation:

$$\begin{aligned}
\bot &\equiv \forall p : *.\, p \\
\neg\varphi &\equiv [\varphi \Rightarrow \bot]
\end{aligned}$$

There are two kinds of abstractions and applications of level 1. We introduce some new notation only for the 'polymorphic' ones:

$$\Lambda\mathcal{X} : \Box.\, c \equiv \lambda\mathcal{X} : \Box.\, c \qquad\qquad \{b\ \mathcal{T}\} \equiv (b\ \mathcal{T})$$

Here b and c are objects and $\mathcal{T}$ is a set.

There are three kinds of abstractions and applications of level 0:

$$\begin{array}{rr} \mathsf{suppose}\ n:\varphi.\,P \equiv \lambda n:\varphi.\,P & [P\ Q] \equiv (P\ Q) \\ \mathsf{let}\ x:\mathcal{S}.\,P \equiv \lambda x:\mathcal{S}.\,P & \langle P\ c\rangle \equiv (P\ c) \\ \mathsf{let}\ \mathcal{X}:\Box.\,P \equiv \lambda\mathcal{X}:\Box.\,P & \langle P\ \mathcal{T}\rangle \equiv (P\ \mathcal{T}) \end{array}$$

Note that for proofs P and Q, the application $[P\ Q]$ corresponds to *modus ponens* in a natural deduction.

3 A Term of Type $\perp$ in λU^-

We consider the following universe:

$$\mathcal{U} \equiv \Pi\mathcal{X}:\Box.\,((\wp\wp\mathcal{X}\to\mathcal{X})\to\wp\wp\mathcal{X})$$

For each term t of type $\wp\wp\mathcal{U}$, we define a term of type $\mathcal{U}$:

$$\tau t \equiv \Lambda\mathcal{X}:\Box.\,\lambda f:(\wp\wp\mathcal{X}\to\mathcal{X}).\,\lambda p:\wp\mathcal{X}.\,(t\ \lambda x:\mathcal{U}.\,(p\ (f\ (\{x\ \mathcal{X}\}\ f))))$$

For each term s of type $\mathcal{U}$, we define a term of type $\wp\wp\mathcal{U}$:

$$\sigma s \equiv (\{s\ \mathcal{U}\}\ \lambda t:\wp\wp\mathcal{U}.\,\tau t)$$

(So we do not consider σ and τ as *terms*.)

We define normal terms of type $\wp\mathcal{U}$ and $\mathcal{U}$, respectively:

$$\begin{aligned} &\Delta \equiv \lambda y:\mathcal{U}.\,\neg\forall p:\wp\mathcal{U}.\,[(\sigma y\ p) \Rightarrow (p\ \tau\sigma y)] \\ &\Omega \equiv \text{the normal form of } \tau\ \lambda p:\wp\mathcal{U}.\,\forall x:\mathcal{U}.\,[(\sigma x\ p) \Rightarrow (p\ x)] \end{aligned}$$

In other words, $\Omega \equiv \Lambda\mathcal{X}:\Box.\,\lambda f:(\wp\wp\mathcal{X}\to\mathcal{X}).\,\lambda p:\wp\mathcal{X}.\,\forall x:\mathcal{U}.\,[(\sigma x\ \lambda y:\mathcal{U}.\,(p\ (f\ (\{y\ \mathcal{X}\}\ f)))) \Rightarrow (p\ (f\ (\{x\ \mathcal{X}\}\ f)))]$.

We claim that the following is a term of type $\perp$ in λU^-:

$$\begin{gathered} [\mathsf{suppose}\ 0:\forall p:\wp\mathcal{U}.\,[\forall x:\mathcal{U}.\,[(\sigma x\ p) \Rightarrow (p\ x)] \Rightarrow (p\ \Omega)]. \\ [[\langle 0\ \Delta\rangle\ \mathsf{let}\ x:\mathcal{U}.\,\mathsf{suppose}\ 2:(\sigma x\ \Delta).\,\mathsf{suppose}\ 3:\forall p:\wp\mathcal{U}.\,[(\sigma x\ p) \Rightarrow (p\ \tau\sigma x)]. \\ [[\langle 3\ \Delta\rangle\ 2]\ \mathsf{let}\ p:\wp\mathcal{U}.\,\langle 3\ \lambda y:\mathcal{U}.\,(p\ \tau\sigma y)\rangle\ ]]\ \mathsf{let}\ p:\wp\mathcal{U}.\,\langle 0\ \lambda y:\mathcal{U}.\,(p\ \tau\sigma y)\rangle] \\ \mathsf{let}\ p:\wp\mathcal{U}.\,\mathsf{suppose}\ 1:\forall x:\mathcal{U}.\,[(\sigma x\ p) \Rightarrow (p\ x)].\,[\langle 1\ \Omega\rangle\ \mathsf{let}\ x:\mathcal{U}.\,\langle 1\ \tau\sigma x\rangle]] \end{gathered}$$

Note that each subterm (except for the term itself) is normal. One easily verifies that (in the empty context) there is no normal term of type $\perp$ in λU^-. At the end of this article, we analyse the β-reduction behaviour of this proof term.

The proof is simple in the sense that it contains just 6 applications corresponding to *modus ponens*. In order to get an idea of the influence of abbreviations, one can also calculate the *length*: the total number of applications, abstractions, products, and occurrences of variables and sorts. For example, the terms abbreviated by $\perp$, $\mathcal{U}$, Δ, and Ω have length 3, 15, 241, and 145. The complete proof term has length 2039.

In order to explain the idea of this proof, we first describe the paradox of Burali-Forti.

4 Burali-Forti's Paradox

Cesare Burali-Forti (1897) published a result that lead to the first paradox in naive set theory. He showed that there are different ordinal numbers α and β such that neither $\alpha < \beta$ nor $\beta < \alpha$, which contradicts a result of Georg Cantor (1897). (In fact, Burali-Forti considered *perfectly ordered* classes instead of well-orderings, so one has to adapt his proof in order to get a contradiction.)

A binary relation $\prec$ on a set $\mathcal{X}$ is called a *well-ordering* if it is connected (for all different x and y in $\mathcal{X}$, $x \prec y$ or $y \prec x$) and well-founded (there is no infinite descending sequence $\ldots \prec x_2 \prec x_1 \prec x_0$ in $\mathcal{X}$). Then it is also irreflexive and transitive. Each member x of $\mathcal{X}$ determines an *initial segment* of $(\mathcal{X}, \prec)$: the set $\{y \in \mathcal{X} | y \prec x\}$, ordered by the restriction of $\prec$ to this set.

An *ordinal number* is the *order type* of a well-ordered set. Let α and β be the order types of the well-ordered sets $(\mathcal{X}, \prec)$ and $(\mathcal{Y}, \prec')$. Then $\beta = \alpha$ expresses that $(\mathcal{Y}, \prec')$ is isomorphic to $(\mathcal{X}, \prec)$ and $\beta < \alpha$ expresses that $(\mathcal{Y}, \prec')$ is isomorphic to an initial segment of $(\mathcal{X}, \prec)$. (This is well-defined, since isomorphic well-ordered sets have isomorphic initial segments.) It is equivalent to the existence of a monotone function from $(\mathcal{Y}, \prec')$ to an initial segment of $(\mathcal{X}, \prec)$.

Assuming that the relation $<$ on the collection $\mathcal{NO}$ of all ordinal numbers is connected, Burali-Forti (could have) showed that it is a well-ordering. So it has an order type Ω.

Let α be the order type of a well-ordered set $(\mathcal{X}, \prec)$. Then the function that assigns to each x in $\mathcal{X}$ the order type of the initial segment of $(\mathcal{X}, \prec)$ determined by x, is an isomorphism from $(\mathcal{X}, \prec)$ to the initial segment of $(\mathcal{NO}, <)$ determined by α. This shows that for each ordinal α, $\alpha < \Omega$. In particular, $\Omega < \Omega$. This contradicts the fact that $<$ is a well-ordering.

4.1 Simplifications of Burali-Forti's Paradox

Burali-Forti's paradox can be simplified in such a way that Cantor's result is irrelevant. Girard (1972) considered the universe $\mathcal{UO}$ of all *orderings without torsion*: irreflexive, transitive relations such that different elements determine non-isomorphic initial segments. The definition of $<$ can be extended to $\mathcal{UO}$. Then the following contradictory statements can be proved in system U:

> An ordering without torsion is not isomorphic to any of its initial segments. $(\mathcal{UO}, <)$ is an ordering without torsion. Each ordering without torsion is isomorphic to an initial segment of $(\mathcal{UO}, <)$.

Coquand (1986) formalised a version by considering the universe of order types of transitive, well-founded relations (and using the definition of $<$ in terms of monotone functions). This version is similar to the paradox of Dimitry Mirimanoff (1917):

> A set x is *well-founded* (with respect to the membership relation) if no infinite descending sequence $\ldots \in x_1 \in x_0 \in x$ exists. The collection $\mathcal{WF}$ of all well-founded sets is well-founded, so $\mathcal{WF} \in \mathcal{WF}$. This contradicts the well-foundedness of $\mathcal{WF}$.

A still simpler paradox is that of Bertrand Russell (1903):

> Let Δ be the collection of all sets x such that $x \notin x$. Then the proposition $\Delta \in \Delta$ is equivalent to its negation.

One could try to formalise this paradox in a type system like λU as follows:

Define some universe $\mathcal{U}$, together with a function σ from $\mathcal{U}$ to its power set $\wp\mathcal{U}$ and a function τ in the other direction, such that for each term X of type $\wp\mathcal{U}$, $(\sigma\ (\tau\ X))$ is β-equal to X. For x and y in $\mathcal{U}$, write $y \in x$ instead of $((\sigma\ x)\ y)$. Write $\{x|x \notin x\}$ instead of $\lambda x : \mathcal{U}. \neg\ x \in x$ and let Δ be the term $(\tau\ \{x|x \notin x\})$. Then the term $\Delta \in \Delta$ of type $*$ is β-equal to its negation. So [suppose $0 : \Delta \in \Delta. [0\ 0]$ suppose $0 : \Delta \in \Delta. [0\ 0]]$ is a proof term of type $\perp$.

However, as noted by Coquand (1986), Russell's paradox cannot be formalised in this way since each proposition has a normal form. (Of course, in an inconsistent system each proposition is *provable* equivalent to its negation.)

4.2 From Ordinal Numbers to Collections of Ordinal Numbers and Back

We return to Burali-Forti's paradox and analyse the connection between $\mathcal{NO}$ and its power set.

For each ordinal number α, let $\sigma\alpha$ be the collection of all smaller ordinals. Let X be a collection of ordinals and let τX be the order type of $(X, \prec)$, where $\prec$ is the restriction of $<$ to X. Then, by definition of $<$, for each ordinal β, $\beta < \tau X$ expresses that β is the order type of some initial segment of $(X, \prec)$. Now assume that for each α in X, all smaller ordinals are also in X. Then each initial segment of $(X, \prec)$ is of the form $(\sigma\alpha, \prec')$ for some α in X, where $\prec'$ is the restriction of $<$ to $\sigma\alpha$. Therefore $\sigma\tau X = \{\beta|\beta < \tau X\} = \{\beta|\beta$ is of the form $\tau\sigma\alpha$ for some α in $X\} = \{\tau\sigma\alpha|\alpha \text{ in } X\}$.

In fact one can show that for each α, $\tau\sigma\alpha = \alpha$, but we will see that we do not need that in order to get a contradiction.

5 Paradoxical Universes

5.1 From a Universe to Its Power Set and Back

Let us call a universe $\mathcal{U}$, together with functions $\sigma : \mathcal{U} \to \wp\mathcal{U}$ and $\tau : \wp\mathcal{U} \to \mathcal{U}$, *paradoxical* if for each X in $\wp\mathcal{U}$, $\sigma\tau X = \{\tau\sigma x|x \text{ in } X\}$.

Each function $f : \mathcal{S} \to \mathcal{T}$ induces a function $f_\star : \wp\mathcal{S} \to \wp\mathcal{T}$ as follows: for each subset X of $\mathcal{S}$, $f_\star X = \{fx|x \text{ in } X\}$. Using this notation, we see that $(\mathcal{U}, \sigma, \tau)$ is paradoxical if and only if the composition $\sigma \circ \tau$ is equal to $(\tau \circ \sigma)_\star$. Note that if $(\mathcal{U}, \sigma, \tau)$ is paradoxical, then $(\wp\mathcal{U}, \sigma_\star, \tau_\star)$ is also paradoxical: $\sigma_\star \circ \tau_\star = (\sigma \circ \tau)_\star = (\tau \circ \sigma)_{\star\star} = (\tau_\star \circ \sigma_\star)_\star$. (Here we need extensionality: if two sets have the same elements, then they are equal.)

5.2 Example of a Paradoxical Universe

Let $\mathcal{U}$ be the universe of all *triples* $(\mathcal{A}, \prec, a)$ consisting of a set $\mathcal{A}$, a binary relation $\prec$ on $\mathcal{A}$, and an element a of $\mathcal{A}$. For each triple $(\mathcal{A}, \prec, a)$, let $\sigma(\mathcal{A}, \prec, a)$ be the collection of all triples of the form $(\mathcal{A}, \prec, b)$, where $b \prec a$. So σ is a function from $\mathcal{U}$ to $\wp\mathcal{U}$. It induces a relation $<$ on $\wp\mathcal{U}$ as follows:

For all collections X and Y of triples, $Y < X$ if and only if Y is in $\sigma_\star X$, that is, if Y is of the form $\sigma(\mathcal{A}, \prec, a)$ for some triple $(\mathcal{A}, \prec, a)$ in X.

For each X in $\wp\mathcal{U}$, let τX denote the triple $(\wp\mathcal{U}, <, X)$.

Now $\sigma\tau X = \sigma(\wp\mathcal{U}, <, X) = \{(\wp\mathcal{U}, <, Y) | Y < X\} = \{\tau Y | Y \text{ is in } \sigma_\star X\} = \{\tau\sigma(\mathcal{A}, \prec, a) | (\mathcal{A}, \prec, a) \text{ in } X\}$.

5.3 Contradiction from the Existence of a Paradoxical Universe

Let $(\mathcal{U}, \sigma, \tau)$ be paradoxical. It is possible to derive a contradiction similar to Russell's paradox:

> Let $\approx$ be the least equivalence relation on $\mathcal{U}$ such that for each x in $\mathcal{U}$, $x \approx \tau\sigma x$. Define a relation $\in$ on $\mathcal{U}$ as follows: $y \in x$ if and only if $y \approx z$ for some z in σx. Let $\Delta \equiv \tau\{x | x \notin x\}$. Prove that for each y in $\mathcal{U}$, $y \in \Delta$ if and only if $y \notin y$. Take $y = \Delta$.

We will derive a contradiction in another way.

Elements of $\mathcal{U}$ will be denoted by $x, y, \ldots$ and subsets of $\mathcal{U}$ by $X, Y, \ldots$.

If y is in σx, then we say that y is a *predecessor* of x and we write $y < x$. Since $(\mathcal{U}, \sigma, \tau)$ is paradoxical, the predecessors of $\tau\sigma x$ are the elements of the form $\tau\sigma y$ for some predecessor y of x (take $X = \sigma x = \{y | y < x\}$). So if $y < x$ then $\tau\sigma y < \tau\sigma x$. (We will use the special case $y = \tau\sigma x$.)

There are several ways to define well-foundedness. The following formulation immediately leads to the principle of *proof by transfinite induction* (without using classical logic or the axiom of choice). Furthermore, the only quantifiers and connectives that it uses are 'for all' and 'if ... then'.

We call X *inductive* if the following holds: for each x, if each predecessor of x is in X, then x itself is in X. We say that x is *well-founded* if x is in each inductive X. (One can easily prove that $\{x | x \text{ is well-founded}\}$ is the *least* inductive subset of $\mathcal{U}$, but we do not use this fact.)

Let $\Omega \equiv \tau\{x | x \text{ is well-founded}\}$. Since $(\mathcal{U}, \sigma, \tau)$ is paradoxical, the predecessors of Ω are of the form $\tau\sigma w$ for some well-founded w.

We claim that Ω is well-founded:

Let X be inductive. In order to show that Ω is in X, we only need to show that each predecessor of Ω is in X. Such a predecessor is of the form $\tau\sigma w$ for some well-founded w. We want to show that w belongs to the set $\{y | \tau\sigma y \text{ is in } X\}$. This follows from the fact that this set is inductive:

Let x be such that for each $y < x$, $\tau\sigma y$ is in X. Then $\tau\sigma x$ is in X since X is inductive and each predecessor of $\tau\sigma x$ is in X, since such a predecessor is of the form $\tau\sigma y$ for some $y < x$.

Note that, until now, we only used the fact that for each X, $\sigma\tau X \subseteq \{\tau\sigma x | x \text{ in } X\}$. Using the other inclusion, we now show that Ω is *not* well-founded:

Suppose that Ω is well-founded. Then $\tau\sigma\Omega$ is of the form $\tau\sigma w$ for some well-founded w, so $\tau\sigma\Omega$ is a predecessor of Ω. On the other hand, $\tau\sigma\Omega \not< \Omega$, since Ω is well-founded and the set $\{y | \tau\sigma y \not< y\}$ is inductive:

Let x be such that for each $y < x$, $\tau\sigma y \not< y$. Then $\tau\sigma x \not< x$. For suppose that $\tau\sigma x < x$. Then $\tau\sigma\tau\sigma x \not< \tau\sigma x$ (take $y = \tau\sigma x$). But $\tau\sigma\tau\sigma x$ is of the form $\tau\sigma y$ for some $y < x$, so $\tau\sigma\tau\sigma x$ is a predecessor of $\tau\sigma x$.

5.4 Formalisation in Pure Type Systems

The preceding derivation of a contradiction from the existence of a paradoxical universe can be formalised in λHOL: we can find a term of type $\bot$ in the context $\mathcal{U} : \Box$, $\sigma : (\mathcal{U} \to \wp\mathcal{U})$, $\tau : (\wp\mathcal{U} \to \mathcal{U})$, $0 : \forall X : \wp\mathcal{U}. (\sigma\ (\tau\ X)) =_{\wp\mathcal{U}} \lambda u : \mathcal{U}. \exists x : \mathcal{U}. ((X\ x) \wedge u =_{\mathcal{U}} (\tau\ (\sigma\ x)))$. Here for each set $\mathcal{A}$, $=_{\mathcal{A}}$ denotes Leibniz equality on $\mathcal{A}$. Instead of $=_{\wp\mathcal{U}}$ one can also take the weaker relation of 'having the same elements'. Since the proof does not use *ex falso sequitur quodlibet* at all, $\bot$ can be replaced by any formula φ.

We need a stronger Pure Type System to prove $\bot$ in the empty context. Let $\mathcal{U}$ be the paradoxical universe given in the example. Using the rule $(\triangle, \Box)$, we formalise the power set $\wp\mathcal{U}$ as the term $\Pi\mathcal{X} : \Box. ((\mathcal{X} \to \wp\mathcal{X}) \to \wp\mathcal{X})$ of type $\Box$. In other words, we read $\Pi u : \mathcal{U}. *$ as abbreviation for $\Pi\mathcal{A} : \Box. \Pi\prec : (\mathcal{A} \to \wp\mathcal{A}). \Pi a : \mathcal{A}. *$. It is not necessary to find a term corresponding to $\mathcal{U}$ itself. For example, $\forall u : \mathcal{U}. (X\ u)$ stands for $\forall\mathcal{A} : \Box. \forall\prec : (\mathcal{A} \to \wp\mathcal{A}). \forall a : \mathcal{A}. ((\{X\ \mathcal{A}\}\ \prec)\ a)$. Note that the rule $(\triangle, *)$ is needed for the quantification over $\Box$. So this can be done in λU.

One can also formalise the preceding paradox in $\lambda\mathrm{U}^-$, using for example the paradoxical universe $\Pi\mathcal{X} : \Box. ((\wp\mathcal{X} \to \mathcal{X}) \to \mathcal{X})$ or the following one:

$$\mathcal{U} \equiv \Pi\mathcal{X} : \Box. ((\wp\mathcal{X} \to \mathcal{X}) \to \wp\mathcal{X})$$

Define a term of type $(\wp\mathcal{U} \to \mathcal{U})$:

$$\tau \equiv \lambda X : \wp\mathcal{U}. \Lambda\mathcal{A} : \Box. \lambda c : (\wp\mathcal{A} \to \mathcal{A}). \lambda a : \mathcal{A}. \varphi$$

Here φ expresses that a is of the form $(c\ (\{x\ \mathcal{A}\}\ c))$ for some $x : \mathcal{U}$ such that $(X\ x)$. (Note that $(\{x\ \mathcal{A}\}\ c) : \wp\mathcal{A}$, so $(c\ (\{x\ \mathcal{A}\}\ c)) : \mathcal{A}$.) This can be done without defining $\exists$, $\wedge$, and $=_{\mathcal{A}}$, as follows:

$$\varphi \equiv \forall P : \wp\mathcal{A}. [\forall x : \mathcal{U}. [(X\ x) \Rightarrow (P\ (c\ (\{x\ \mathcal{A}\}\ c)))] \Rightarrow (P\ a)]$$

Define a term of type $(\mathcal{U} \to \wp\mathcal{U})$:

$$\sigma \equiv \lambda x : \mathcal{U}. (\{x\ \mathcal{U}\}\ \tau)$$

Then one easily verifies that $(\mathcal{U}, \sigma, \tau)$ is paradoxical. In fact, for each X of type $\wp\mathcal{U}$, $(\sigma\ (\tau\ X))$ is β-equal to the term corresponding to "the intersection of all subsets P of $\mathcal{U}$ containing $(\tau\ (\sigma\ x))$ for each x in X". This simplifies the formal proof term, since β-conversion between two propositions φ and χ does not "count" as a proof step: if P is a proof term of type φ, then P also has type χ.

In this way, one finds a term of type $\perp$ in λU^- that uses *modus ponens* 12 times. It is of the form $[P\ Q]$, where P is a normal term of type "Ω is not well-founded" and Q is a normal term of type "Ω is well-founded". The terms $\mathcal{U}$, Ω, "Ω is well-founded", P, and Q have length 11, 163, 285, 1849, and 1405.

6 Powerful Universes

The proof term that we presented earlier, is shorter and has a simpler reduction behaviour. Furthermore, we defined terms τt and σs without using quantifiers or connectives. The main idea of the simplification is to consider the power set of the power set of some universe $\mathcal{U}$. In fact, we already considered $\wp\wp\mathcal{U}$ implicitly: Let for each subset C of $\wp\mathcal{U}$, $\bigcap C$ be the *intersection* of all members Y of C, that is, $\bigcap C \equiv \{y|\text{for each } Y \text{ in } C,\ y \text{ is in } Y\}$. Then, by definition, $\{x|x \text{ is well-founded}\} \equiv \bigcap\{X|X \text{ is inductive}\}$ and for each X in $\wp\mathcal{U}$, $\{\tau\sigma x|x \text{ in } X\} \equiv \bigcap\{Y|\text{for each } x \text{ in } X,\ \tau\sigma x \text{ is in } Y\}$. In the example of a paradoxical universe, we defined $\sigma(\mathcal{A}, \prec, a) \equiv \bigcap\{X|\text{for each } b \prec a,\ (\mathcal{A}, \prec, b) \text{ is in } X\}$. The relation $\prec$ on $\mathcal{A}$ induces a function $s : \mathcal{A} \to \wp\wp\mathcal{A}$ as follows: $sa = \{B \text{ in } \wp\mathcal{A}|\text{for each } b \prec a,\ b \text{ is in } B\}$. In terms of this function, $\sigma(\mathcal{A}, \prec, a) = \bigcap\{X|\{b \text{ in } \mathcal{A}|(\mathcal{A}, \prec, b) \text{ is in } X\} \text{ is in } sa\}$. Note that if $\prec$ is (Leibniz) equality on $\mathcal{A}$, then the function s can be defined without using quantifiers or connectives: $sa = \{B \text{ in } \wp\mathcal{A}|a \text{ is in } B\}$.

By using the fact that no set is isomorphic to the power set of its power set, John Reynolds (1984) proved that there is no set-theoretic model of polymorphic (or second-order) typed λ-calculus. By refining this result and using a computer, Coquand (1994) found a formal proof of a contradiction in system U^-. He considered the universe $\mathsf{A}_0 \equiv \Pi\mathcal{X} : \Box.\ ((\wp\wp\mathcal{X}\to\mathcal{X})\to\mathcal{X})$ and defined functions $\mathsf{match} : \mathsf{A}_0 \to \wp\wp\mathsf{A}_0$ and $\mathsf{intro} : \wp\wp\mathsf{A}_0 \to \mathsf{A}_0$. Then he showed that these functions constitute an isomorphism with respect to certain partial equivalence relations. In fact, $(\mathsf{A}_0, \mathsf{match}, \mathsf{intro})$ can also be used to formulate a Burali-Forti like paradox: it is an example of a *powerful universe.*

6.1 From a Universe to the Power Set of Its Power Set and Back

Let us call a universe $\mathcal{U}$, together with functions $\sigma : \mathcal{U} \to \wp\wp\mathcal{U}$ and $\tau : \wp\wp\mathcal{U} \to \mathcal{U}$, *powerful* if for each C in $\wp\wp\mathcal{U}$, $\sigma\tau C = \{X|\text{the set } \{y|\tau\sigma y \text{ is in } X\} \text{ is in } C\}$.

Each function $f : \mathcal{S} \to \mathcal{T}$ induces a function $f^\star : \wp\mathcal{T} \to \wp\mathcal{S}$ as follows: for each subset Y of $\mathcal{T}$, $f^\star Y = \{x \text{ in } \mathcal{S}|fx \text{ is in } X\}$. Using this notation, we see that $(\mathcal{U}, \sigma, \tau)$ is powerful if and only if the composition $\sigma \circ \tau$ is equal to $(\tau \circ \sigma)^{\star\star}$. Note that if $(\mathcal{U}, \sigma, \tau)$ is powerful, then $(\wp\mathcal{U}, \tau^\star, \sigma^\star)$ is also powerful: $\tau^\star\circ\sigma^\star = (\sigma\circ\tau)^\star = (\tau\circ\sigma)^{\star\star\star} = (\sigma^\star\circ\tau^\star)^{\star\star}$. (Here we do not need extensionality.)

6.2 Example of a Powerful Universe

Let $\mathcal{U}$ be the universe of all *triples* $(\mathcal{A}, s, a)$ consisting of a set $\mathcal{A}$, a function $s : \mathcal{A} \to \wp\wp\mathcal{A}$, and an element a of $\mathcal{A}$. For each triple $(\mathcal{A}, s, a)$, let $\sigma(\mathcal{A}, s, a)$ be the collection of all subsets X of $\mathcal{U}$ such that $\{b \text{ in } \mathcal{A} | (\mathcal{A}, s, b) \text{ is in } X\}$ is in sa. Since σ is a function from $\mathcal{U}$ to $\wp\wp\mathcal{U}$, $\sigma^{\star\star}$ is a function from $\wp\wp\mathcal{U}$ to $\wp\wp\wp\wp\mathcal{U}$. For each C in $\wp\wp\mathcal{U}$, let τC denote the triple $(\wp\wp\mathcal{U}, \sigma^{\star\star}, C)$.

In order to verify that $(\mathcal{U}, \sigma, \tau)$ is powerful, let C in $\wp\wp\mathcal{U}$ and X in $\wp\mathcal{U}$. Then the following propositions are equivalent (by definition):

X is in $\sigma\tau C$;
X is in $\sigma(\wp\wp\mathcal{U}, \sigma^{\star\star}, C)$;
$\{b \text{ in } \wp\wp\mathcal{U} | (\wp\wp\mathcal{U}, \sigma^{\star\star}, b) \text{ is in } X\}$ is in $\sigma^{\star\star}C$;
$\sigma^\star\{b \text{ in } \wp\wp\mathcal{U} | \tau b \text{ is in } X\}$ is in C;
$\{y \text{ in } \mathcal{U} | \tau\sigma y \text{ is in } X\}$ is in C.

6.3 Contradiction from the Existence of a Powerful Universe

Let $(\mathcal{U}, \sigma, \tau)$ be powerful. We will derive a contradiction in a similar way as for paradoxical universes.

Elements of $\mathcal{U}$ will be denoted by $x, y, \ldots$ and subsets of $\mathcal{U}$ by $X, Y, \ldots$. For each x, σx is in $\wp\wp\mathcal{U}$. $(\mathcal{U}, \sigma, \tau)$ is powerful, so:

$$\sigma\tau\sigma x = \{X | \text{the set } \{y | \tau\sigma y \text{ is in } X\} \text{ is in } \sigma x\}$$

We say that y is a *predecessor* of x (and we write $y < x$) if for each X in σx, y is in X (in other words, if y is in $\bigcap \sigma x$). One can easily prove that if $y < x$ then $\tau\sigma y < \tau\sigma x$. We will only do this for the special case $y = \tau\sigma x$. Note that if X is in σx, then each predecessor of x is in X.

X is called *inductive* if the following holds: for each x in $\mathcal{U}$, if X is in σx, then x is in X. We say that x is *well-founded* if x is in each inductive X. (Note that it is *not* clear whether $\{x | x \text{ is well-founded}\}$ is inductive: if one tries to prove this, one would like to use something like: if Y is in σx and $Y \subseteq X$, then X is in σx.)

Let $\Omega \equiv \tau\{X | X \text{ is inductive}\}$. $(\mathcal{U}, \sigma, \tau)$ is powerful, so:

$$\sigma\Omega = \{X | \text{ the set } \{y | \tau\sigma y \text{ is in } X\} \text{ is inductive}\}$$

We claim that Ω is well-founded:

Let X be inductive. In order to prove that Ω is in X, we only need to show that X is in $\sigma\Omega$. In other words, we show that the set $\{y | \tau\sigma y \text{ is in } X\}$ is inductive. So let x be in $\mathcal{U}$. Since X is inductive, we have the following: if X is in $\sigma\tau\sigma x$, then $\tau\sigma x$ is in X. In other words, if the set $\{y | \tau\sigma y \text{ is in } X\}$ is in σx, then x is in $\{y | \tau\sigma y \text{ is in } X\}$. This is exactly what we had to prove.

In order to show that Ω is *not* well-founded, we first prove that the set $\{y | \tau\sigma y \not< y\}$ is inductive:

Let x be such that $\{y | \tau\sigma y \not< y\}$ is in σx. Then $\tau\sigma x \not< x$. For suppose that $\tau\sigma x < x$. In other words, for each X in σx, $\tau\sigma x$ is in X. Applying this to the set $\{y | \tau\sigma y \not< y\}$, which is in σx, we see that $\tau\sigma\tau\sigma x \not< \tau\sigma x$. On the other hand,

$\tau\sigma\tau\sigma x < \tau\sigma x$: Let X be in $\wp\mathcal{U}$. We have to show the following: if X is in $\sigma\tau\sigma x$, then $\tau\sigma\tau\sigma x$ is in X. In other words, if $\{y|\tau\sigma y$ is in $X\}$ is in σx, then $\tau\sigma x$ is in $\{y|\tau\sigma y$ is in $X\}$. This follows from the assumption that $\tau\sigma x < x$, i.e. for each Y in σx, $\tau\sigma x$ is in Y.

Now suppose that Ω is well-founded. Then, since $\{y|\tau\sigma y \not< y\}$ is inductive, $\tau\sigma\Omega \not< \Omega$. On the other hand, $\tau\sigma\Omega < \Omega$: Let X be in $\wp\mathcal{U}$. We have to show: if X is in $\sigma\Omega$, then $\tau\sigma\Omega$ is in X. In other words, if the set $\{y|\tau\sigma y$ is in $X\}$ is inductive, then Ω is in $\{y|\tau\sigma y$ is in $X\}$. This follows from the assumption that Ω is well-founded, i.e. for each inductive Y, Ω is in Y.

7 Reduction Behaviour

Douglas Howe (1987) used a computer to study the reduction behaviour of a massive term corresponding to one particular proof of Girard's paradox. Just like the proofs we gave, it did not use *ex falso sequitur quodlibet*, so $\bot$ can be replaced by a variable of type $*$. Using this, Howe constructed a *looping combinator* (but not a fixed-point combinator). (See (Coquand and Herbelin 1994) and also (Geuvers and Werner 1994).)

We now return to the proof term that we presented in Section 3. It formalises the preceding derivation of a contradiction, using some other powerful universe than the two that we mentioned earlier.

One easily verifies that $(\mathcal{U}, \lambda s : \mathcal{U}.\, \sigma s, \lambda t : \wp\wp\mathcal{U}.\, \tau t)$ is powerful: in fact, for each term t of type $\wp\wp\mathcal{U}$, the term $\sigma\tau t$ β-reduces to $\lambda p : \wp\mathcal{U}.\, (t\ \lambda x : \mathcal{U}.\, (p\ \tau\sigma x))$. One can calculate the normal form of $(\{\tau\sigma\tau\sigma \cdots \tau\sigma x\ \mathcal{X}\}\ f)$, where x, $\mathcal{X}$, and f are variables. It contains nested expressions of the form $\lambda p : \wp\mathcal{X}.\, (\sigma x'\ \lambda x'' : \mathcal{U}.\, (p\ (f\ \cdots)))$.

For each term s of type $\mathcal{U}$, let $\Theta^0 s$ be s and, for each natural number n, let $\Theta^{n+1} s$ be the normal form of $\tau\sigma\Theta^n s$. For each term p of type $\wp\mathcal{U}$, let $\Theta_0^\star p$ be p and, for each natural number n, let $\Theta_{n+1}^\star p$ be $\lambda y : \mathcal{U}.\, (p\ \Theta^{n+1} y)$. Then, for variables x and p, for each natural number n, the normal form of $(\sigma\Theta^n x\ p)$ is $(\sigma x\ \Theta_n^\star p)$.

The fact that β-reduction of the proof term goes on indefinitely, is caused by steps that correspond to the rule $(*, *)$, that is, replacing a subterm of the form $[\mathsf{suppose}\ n : \varphi.\, P\ Q]$ by $P[Q/n]$. One can show that each infinite sequence of β-reduction steps, starting with a term in λU, contains such a step. So we can concentrate on "big steps": steps that correspond to $(*, *)$, followed by a maximal sequence of steps corresponding to other rules.

Let n be a natural number. We first define two propositions:

$$\varphi_n \equiv \forall p : \wp\mathcal{U}.\, [\forall x : \mathcal{U}.\, [(\sigma x\ \Theta_n^\star p) \Rightarrow (p\ \Theta^n x)] \Rightarrow (p\ \Theta^n \Omega)]$$
$$\psi_n \equiv \forall x : \mathcal{U}.\, [(\sigma x\ \Theta_n^\star \Delta) \Rightarrow \neg\forall p : \wp\mathcal{U}.\, [(\sigma x\ \Theta_n^\star p) \Rightarrow (p\ \Theta^{n+1} x)]]$$

So φ_n expresses that $\Theta^n\Omega$ is in each subset X of $\mathcal{U}$ for which $\{y|\Theta^n y$ is in $X\}$ is inductive. Note that φ_0 is "Ω is well-founded" and φ_{n+1} is the normal form of "$\Theta^{n+1}\Omega < \Theta^n\Omega$". The proposition ψ_n expresses that $\{y|\Theta^{n+1}y \not< \Theta^n y\}$ is inductive. We also define five proof terms:

$$\begin{aligned}
R_n &\equiv \mathsf{let}\ p : \wp\mathcal{U}.\,\mathsf{suppose}\ 1 : \forall x : \mathcal{U}.\,[(\sigma x\ \Theta^\star_n p) \Rightarrow (p\ \Theta^n x)].\\
&\qquad [\langle 1\ \Omega\rangle\ \mathsf{let}\ x : \mathcal{U}.\,\langle 1\ \tau\sigma x\rangle]\\
M_n &\equiv \mathsf{let}\ x : \mathcal{U}.\,\mathsf{suppose}\ 2 : (\sigma x\ \Theta^\star_n \Delta).\\
&\qquad \mathsf{suppose}\ 3 : \forall p : \wp\mathcal{U}.\,[(\sigma x\ \Theta^\star_n p) \Rightarrow (p\ \Theta^{n+1} x)].\\
&\qquad [[\langle 3\ \Delta\rangle\ 2]\ \mathsf{let}\ p : \wp\mathcal{U}.\,\langle 3\ \lambda y : \mathcal{U}.\,(p\ \tau\sigma y)\rangle]\\
P_n &\equiv \mathsf{suppose}\ 4 : \psi_n.\,\mathsf{suppose}\ 0 : \varphi_n.\,[[\langle 0\ \Delta\rangle\ 4]\ \mathsf{let}\ p : \wp\mathcal{U}.\,\langle 0\ \lambda y : \mathcal{U}.\,(p\ \tau\sigma y)\rangle]\\
L_n &\equiv \mathsf{suppose}\ 0 : \varphi_n.\,[[\langle 0\ \Delta\rangle\ M_n]\ \mathsf{let}\ p : \wp\mathcal{U}.\,\langle 0\ \lambda y : \mathcal{U}.\,(p\ \tau\sigma y)\rangle]\\
Q_n &\equiv \mathsf{suppose}\ 4 : \psi_n.\,[\langle 4\ \Omega\rangle\ \mathsf{let}\ x : \mathcal{U}.\,\langle 4\ \tau\sigma x\rangle]
\end{aligned}$$

Then R_n proves φ_n, M_n proves ψ_n, P_n proves $[\psi_n \Rightarrow \neg\varphi_n]$, L_n proves $\neg\varphi_n$, and Q_n proves $[\psi_n \Rightarrow \neg\varphi_{n+1}]$.

Note that $[L_0\ R_0]$ is the proof term that we presented in Section 3.

For each natural number n, $[[P_n\ M_n]\ R_n]$ reduces in one step to $[L_n\ R_n]$. (Variable 4 disappears.) This reduces in a big step to $[[Q_n\ M_n]\ R_{n+1}]$. (Variable 0 disappears and some occurrences of 1 are renamed as 4.) This reduces to $[[P_{n+1}\ M_{n+1}]\ R_{n+1}]$. (Variable 4 disappears and some occurrences of 2 and 3 are renamed as 4 and 0.)

So these proof terms of type $\perp$ in $\lambda\mathrm{U}^-$ reduce in three big steps to a similar proof term: only the types of the proof variables change a little bit.

References

Barendregt, H.P.: Typed lambda calculi, in: *Handbook of Logic in Computer Science (Vol. 2)*, S. Abramsky et al. (editors), Clarendon Press, Oxford (1992)

Burali-Forti, C.: Una questione sui numeri transfiniti, *Rendiconti del Circolo Matematico di Palermo* **11** (1897) 154–164

Cantor, G.: Beiträge zur Begründung der transfiniten Mengenlehre, II, *Mathematische Annalen* **49** (1897) 207–246

Coquand, Th.: An analysis of Girard's paradox, in: *Proceedings Symposium on Logic in Computer Science: Cambridge, Massachusetts, June 16–18, 1986*, IEEE Computer Society Press, Washington, D.C. (1986) 227–236

Coquand, Th.: A New Paradox in Type Theory, in: *Logic and philosophy of science in Uppsala: papers from the 9th international congress of logic, methodology and philosophy of science*, D. Prawitz, D. Westerstaahl (editors), Kluwer Academic Publishers, Dordrecht (1994) ?–?

Coquand, Th., Herbelin, H.: *A*-translation and looping combinators in pure type systems, *Journal of Functional Programming* **4** (1994) 77–88

Geuvers, J.H.: *Logics and Type Systems*, Proefschrift, Katholieke Universiteit Nijmegen (1993)

Geuvers, H., Werner, B.: On the Church-Rosser property for Expressive Type Systems and its Consequences for their Metatheoretic Study, in: *Proceedings of the Ninth Annual Symposium on Logic in Computer Science, Paris, France*, IEEE Computer Society Press, Washington, D.C. (1994) 320–329

Girard, J.-Y.: *Interprétation fonctionnelle et élimination des coupures de l'arithmétique d'ordre supérieur*, Thèse de Doctorat d'État, Université Paris VII (1972)

Howe, D.J.: The Computational Behaviour of Girard's Paradox, in: *Proceedings Symposium on Logic in Computer Science: Ithaka, New York, June 22–25, 1987*, IEEE Computer Society Press, Washington, D.C. (1987) 205–214

Mirimanoff, D.: Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles, *L'Enseignement Mathématique* **19** (1917) 37–52

Reynolds, J.C.: Polymorphism is not Set-Theoretic, in: *Semantics of Data Types*, G. Kahn et al. (editors), Lecture Notes in Computer Science **173**, Springer-Verlag, Berlin Heidelberg (1984) 145–156

Russell, B.: *The Principles of Mathematics*, Cambridge University Press, Cambridge, G.B. (1903)