Philippe de Rouilhan

Towards finishing off the axiom of reducibility

Philosophia Scientiæ, tome 1, nº 3 (1996), p. 17-35
<http://www.numdam.org/item?id=PHSC_1996__1_3_17_0>

Towards Finishing off the Axiom of Reducibility

Philippe de Rouilhan

Institut d'Histoire et Philosophie des Sciences et des Techniques

Philosophia Scientiæ, 1 (3), 1996, 17-35.

Philippe de Rouilhan

Abstract. This article* is about Russell's theory of types and, more precisely, about the axiom of reducibility. Since this axiom appeared, none of the criticisms it has been subjected to by Russell himself, then by Poincaré, Wittgenstein, Chwistek, etc., invalidates it except, it seems, that of Gödel [1944]. But Gödel's criticism is informal and dogmatic. I propose a formalization of this criticism and an argument from analogy in favor of it; along the way, I refute the criticism Charles Parsons has endeavored to make of Gödel criticism in volume II of Gödel's *Collected Works*. This, of course, presupposes the prior formalisation of Russell's theory of types itself.

Résumé. Cet article porte sur la théorie russellienne des types, et plus précisément sur l'axiome de réductibilité. Aucune des critiques dont il a été l'objet de la part de Russell lui-même, puis de Poincaré, de Wittgenstein, de Chwistek, etc., depuis son apparition n'est dirimante sauf, semble-t-il, celle de Gödel [1944]. Mais la critique de Gödel reste informelle et dogmatique. Je proposerai une formalisation de cette critique et un argument analogique en sa faveur ; et je réfuterai, au passage, la critique que, dans le volume II des *Collected Works* de Gödel, Ch. Parsons croit pouvoir faire de la critique gödélienne. Cela suppose, évidemment, une formalisation préalable de la théorie russellienne des types elle-même.

1. The Vicious Circle Principle, the Notion of Order

Most scholars have emphasized the 'appalling complexity', the 'overwhelming proliferation' involved in the Russellian hierarchy of types, its 'labyrinthine' character. It must have been of the 'beastly theory of types', as Wittgenstein [1913a] termed it, that Russell was at times thinking when he cried out in the course of their conversations: "Logic's hell!" [Wittgenstein 1977, 30]. However, when it comes to the theory of types, it is Russell's approach, more than the thing itself, which is beastly. In his article on Russell's logic, Gödel decried the lack of 'formal precision' of *Principia Mathematica*, and the absence, "above all, [of] a precise statement of the syntax of formalism" [Gödel 1944, 126]. The work of logicians, that of Schütte [1960, §27] or of Church [1976], for example, has shown well that this *kind* of theory can be mastered on the formal level. The problem has remained, however, of mastering *Russell*'s theory, and not just the 'syntax of formalism', as Gödel wished, but its intended interpretation as well.

---

* Before being presented on May 17, 1994 at the *Congrès International Henri Poincaré, Nancy, May 14-18, 1994*, the content of this paper was first presented at the May 3, 1994 meeting of the *Séminaire d'épistémologie comparative*, Aix-en-Provence, upon the invitation of Prof. E. Schwartz. A much more detailed presentation is to be found in my *Russell et le cercle des paradoxes* [Rouilhan 1996], especially chap. VI, §§ 11-14 and 18-23]. I thank C. Hill for her help in putting the present version into English.

The theory I will present is the result of a certain compromise made regarding some not very easily reconcilable demands: that of remaining historically faithful to Russell, but also that of theoretical consistency, and yet again that of placing Russell's theory with a historical perspective which makes it intelligible. This theory will be a theory of individuals, propositions, and (propositional) functions. It will satisfy the vicious circle principle in the technical form Russell gave it in 1908:

> Whatever contains an apparent variable must not be a possible value of that variable. [Russell 1908, 237; 1956, 75][1]

But I will extend the idea of apparent variable further than Russell did, extending it to the variables which stand for possible arguments in a function and, in Russell's notation, are sometimes indicated by a circumflex. This will allow me to integrate into the vicious circle principle, as Russell understood it, the principle regarding functions he was led to add to it.

Putting the vicious circle principle into use involves the idea of *order*. Let us, as Russell sometimes did in heuristic presentations of the vicious circle principle, talk in terms of 'presupposition'. An entity 'presupposes' the values of its apparent variables. This is why it cannot be one of these values. It also 'presupposes' the values of the apparent variables these values may contain and cannot be one of these new values either. Etc. The idea is to assign an 'order' $n \in \mathrm{N}$ to each entity, measuring, so to speak, its 'depth of presupposition', and to comply with the vicious circle principle by having apparent variables take values of orders strictly lower than those of the entities in which they occur. In particular, the arguments of a function will be of strictly lower orders than that of the function. Moreover, to remain faithful to Russell, not only will the orders of the arguments have an *upper bound* . For each place of the argument, the order will also be *determinate*. One is finally led to define the order of an entity as follows: the order of an individual will be 0; the order of a proposition or function will be the least whole number greater than or equal to the orders of its constants (in the extra-linguistic sense), and strictly greater than the orders (of the values) of its variables. As concerns language, the order of a symbol for an entity, or entities (if this symbol contains real variables), will be that of the corresponding entities, and this will be the least whole number greater than or equal

1 The notion of apparent (bound) variable must be understood here, with Russell, in an extra-ordinary, extra-linguistic sense. An entity contains an apparent variable in this sense if its logically perfect expression contains one in the ordinary, linguistic sense. The notion of real (free) variable is obviously not capable of playing this dual role.

to the orders of its constants and of its real variables, and strictly greater than those of its apparent variables.

2. The Notion of Type, the Hierarchy of Types

As for principles for the Russellian hierarchy of individuals, propositions and functions, there was not only the vicious circle principle, but also a principle analogous to the one which had led Frege to his own hierarchy of functions. The notion of order was only one of the notions involved in the notion of '*type*'. Trying to remain faithful, in all essentials, to Russell's thought, one is led to assign to individuals a certain type, say ι; to propositions of the n^{th} order a certain type determined by this order alone, say π_n; and to a function of the n^{th} order a certain type determined by this order and the types $r_1, ..., r_k$ of its arguments, say $\langle n, r_1, ..., r_k \rangle$.

The set T_r of the *types* in question may be defined as the smallest set X with a function of order od: $X \rightarrow N$ such that (for all $n \geq 0$, all $k \geq 1$, and all $r_1, ..., r_k$)

(i) $\iota, \pi_n \in X$; $\text{od}(\iota) = 0$, $\text{od}(\pi_n) = n$;

(ii) $(r_1, ..., r_k \in X \text{ and } n \geq 1 + \max_{1 \leq i \leq k} \text{od}(r_i)) \Rightarrow$
$(\langle n, r_1, ..., r_k \rangle \in X \text{ and } \text{od}(\langle n, r_1, ..., r_k \rangle) = n)$.

Each type r, therefore, has its order od(r), and one has:

$$\text{od}(\iota) = 0,$$
$$\text{od}(\pi_n) = n,$$
$$\text{od}(\langle n, r_1, ..., r_k \rangle) = n \geq 1 + \max_{1 \leq i \leq k} \text{od}(r_i)).$$

The types can be compared in terms of their orders, and a structure of complete quasi-order given to the set T_r in this way. This is what is called the hierarchy of types.

Naturally, the so-called *theory of types* is not the theory having as its objects the types just discussed, but the theory of individuals, propositions and functions set out in accordance with the hierarchy of types just discussed.

3. Towards the Formalization of the Theory of Types: the Language L(TTR)

L(TTR) will be a language with infinitely many sorts of variables, namely, for every $r \in \mathrm{T_r}$, variables 'x_1^r', 'x_2^r', 'x_3^r', ... of entities of type r: variables of individuals if r = i, variables of propositions if r is of the form p_n, and variables of functions if r is of the form $\langle n, r_1, ..., r_k \rangle$. Besides the variables may occur constants, namely, for any $r \in \mathrm{T_r}$, the constants 'a_1^r', 'a_2^r', 'a_3^r' ... for entities of type r: constants of individual if $r = \iota$, constants of proposition if r is of the form π_n, and constants of function if r is of the form $\langle n, r_1, ..., r_k \rangle$[2].

Among the primitive symbols are also to be found the logical symbols '$\neg$', '$\Rightarrow$', and '$\forall$', corresponding (up to the typological character of the system) to ordinary negation, (material) implication, and universal quantification respectively; the symbol 'λ', corresponding to functional abstraction; and parentheses for punctuation.

The well-formed symbols (or well-formed expressions) will be defined recursively, and, except for the variables and constants, will be, *grosso modo*, of one of the following forms: '$\neg p$', '$p \Rightarrow q$', '$(\forall x)p$', '$(\lambda xy ...)p$', 'fxy ...'. At the same time the well-formed symbols are defined, they will be assigned a type, and the notion of the real (resp. apparent) occurrence of a variable in such a symbol will be introduced. For whatever values (of the appropriate type) assigned to the variables for the real occurrences they have there, a well-formed symbol of type r will correspond to an entity of type r (individual, proposition, or function, depending on the case), and, for this reason, will be called an 'entity symbol', or a 'complete symbol'. It will be more specifically called a 'formula' if these entities are propositions, and a 'term' if not. A complete symbol without any occurrence of a variable will be said to be 'closed', corresponding to a definite entity. A closed formula will be called a 'sentence'. Now, here is the

Rigorous Definition of the Language L(TTR)

1- *Alphabet*

1.1- The variables 'x_n^r', where $r \in \mathrm{T_r}$ and $n \geq 1$

2 The abuse of language in these explanations will not have gone unnoticed: it will never be the letter 'r' itself which occurs as a superscript of a variable or a constant, but the canonical name of a certain $r \in \mathrm{T_r}$.

1.2- Possibly, constants ‘ a_n^r ’, where $r \in \mathrm{T_r}$ and $n \geq 1$

1.3- ‘$\neg$’ , ‘$\Rightarrow$’ , ‘$\forall$’ , ‘λ’ , ‘(‘ , ‘)’

2- *Complete symbols, formulas, terms, closed formulas (sentences), and closed terms*

2.1- The variables and constants with a superscript r are complete symbols of type r. The occurrence of a variable is real in the complete symbol it constitutes.

2.2- If c_1 and c_2 are complete symbols of types π_n, $\pi_{n'}$ respectively, then

$$\ulcorner(\neg c_1)\urcorner \text{ and } \ulcorner(c_1 \Rightarrow c_2)\urcorner$$

are complete symbols of types π_n, $\pi_{n''}$ respectively, where $n'' = \max(n, n')$. The occurrences of variables which were real (resp. apparent) in c_1 or c_2 remain so in the new complete symbols.

2.3- If c is a complete symbol of type π_n, and v a variable of type r having a real occurrence in c, then

$$\ulcorner(\forall v)c\urcorner$$

is a complete symbol of type $\pi_{n'}$, where $n' = \max(n, 1 + \mathrm{od}(r))$. The occurrences of variables which were real (resp. apparent) in c remain so in the new complete symbol, except for the occurrences of v which were real, which are now apparent.

2.4- If c is a complete symbol of type π_n, and $v_1, ..., v_k$ variables of types $r_1, ..., r_k$ respectively having real occurrences in c, then

$$\ulcorner(\lambda v_1 \ldots v_k)c\urcorner$$

is a complete symbol of type $\langle n', r_1, ..., r_k\rangle$, where $n' = \max(n, 1 + \max_{1\leq i\leq k}\mathrm{od}(r_i))$. The occurrences of variables which were real (resp. apparent) in c remain so in the new complete symbol, except for the occurrences of $v_1, ..., v_k$ which were real, which are now apparent.

2.5- If $c, c_1, ..., c_k$ are complete symbols of types $\langle n, r_1, ..., r_k\rangle, r_1, ..., r_k$, then

$$\ulcorner c\ c_1 \ldots c_k\urcorner$$

is a complete symbol of type π_n. The occurrences of variables which were real (resp. apparent) in c or in one of the c_i remain so in the new complete symbol.

2.6- A complete symbol is a formula or a term depending on whether its type is or is not of the form π_n.

2.7- A complete symbol without any real occurrence of a variable is closed; a closed formula is a sentence.

The system **TTR** would theoretically be obtainable from language L(**TTR**) by adding axioms and rules of inference corresponding to its intended interpretation. Indeed, besides the axioms and rules of inference obviously expected, Russell introduced certain problematical axioms — the axiom of reducibility, the axiom of infinity, and the axiom of choice — into his own theory of types with the intent of making it powerful enough for the whole of mathematics to be developed within it, and so for verifying the 'logicist' thesis.

4. The Intented Interpretation of the Langage L(TTR)

Let us give a more precise description of the intended interpretation of language L(**TTR**) as Russell himself might have, a more precise description of the hierarchy of individuals, propositions, and functions constituting this interpretation. The two features of this hierarchy it is appropriate to emphasize are as follows.

First, at the bottom of the hierarchy are to be found the 'individuals'. But they are all worthy of the name only in the typological sense. For among them are to be found not only *things* (individuals in the ontological sense, for example Paul and Mary), but also *concepts* (for example the predicate *rides*, the binary relation *loves*)[3]. It is through the only operation of *predication*[4] and, possibly,

3 1°) It was during the *Principles* period that Russell spoke of 'things' and 'concepts' ('predicates' or 'relations'); during the *Principia* period, he spoke rather of 'particulars' and 'universals', and the change was not just terminological. But that is unimportant for the matters which concern us here. 2°) Reading Russell, Gödel seems to have had in mind a hierarchy in which the individuals in the typological sense coincide with the individuals in the ontological sense. Genetic considerations could shed some light on the reason why this not so in Russell's own work [cf. Rouilhan 1996].

4 Russell does not impose this terminology, which I use here for want of better.

that of propositional connection, that the propositions without variables — propositions of order 0 — are constructed (for exemple *Paul rides, Paul loves Mary*).

Second, beyond the individuals, the hierarchy of propositions and functions is built up step by step through the operations of predication, connection, quantification, functional abstraction, and application of a function to arguments, *without any additional ontological data*. In *this* sense, which is not at all anti-realist, one can say that the hierarchy is '*constructive*'. There is nothing mysterious about the reason for this constructivity. On the contrary, what would be mysterious, from a Russellian point of view — and this is understandable — would be a proposition or a function which would not be constructed in this sense. Think about the intended model of the ordinary simple extensional theory of types, about the hierarchy of individuals, classes of individuals, binary (resp. ternary, etc.) relations between individuals, classes of classes of individuals, etc., etc. This hierarchy too has, in an analogous sense, an obvious constructive character, and, except for questioning the typological character of the theory, no one would think of a class or relation not constructible in this sense. Again, think about the intended model of Zermelo's set theory [Zermelo 1908, but also and above all Mirimanoff 1917 and Zermelo 1930], about what is called (deceptively so, since there is nothing typological about the theory in question) the 'cumulative hierarchy of types'. The members of this hierarchy are the individuals (*Urelemente*) and the sets constructible out of them through the iterated application (extending into the transfinite) of operations on individuals or sets which have been already constructed (separation, union, etc.). In Mirimanoff's terminology, these sets are the 'ordinary' sets, the others would be 'extraordinary'[5]. Russellian constructivism is not essentially different.

One can fairly easily picture the Russellian hierarchy in the following way. One first of all defines the language L*(**TTR**), independently of any requirement regarding effectivness. For that, one supplies oneself with a set of constants 'a_i^{ι}' in a one-to-one correspondence with the individuals supposed to be given at the lowest level of the hierarchy — and, for each of them, an 'arity': 0 if it corresponds to a thing, 1 if it corresponds to a predicate, $n \geq 2$

The predication in question is not to be confused with the application of a function to arguments.

5 Indeed, Mirimanoff did not exclude extraordinary sets, but his work was not formalized. For a formal theory including explicitly such sets, see [Aczel 1987].

if it corresponds to a n-ary relation. *Literally*, the definition of L*(**TTR**) differs from that of L(**TTR**) in only the two following ways. On the one hand, the constants of L*(**TTR**) are the ' a_i^{ι} ', and there are no others (see below, footnote 6, rule 2.1-a); on the other hand, a new rule assures that one may construct complete symbols of type π_0 using these constants or variables of type ι (see below, footnote 6, rule 2.1-b)[6].

---

6 Here is the *definition* in extenso *of the language* L*(**TTR**):

1- *Alphabet*

1.1- The variables ' x_n^r ', where $r \in \mathrm{T_r}$ and $n \geq 1$

1.2- The constants ' a_i^{ι} ', with their respective arities $n \geq 0$

1.3- '¬', '⇒', '∀', 'λ', '(', ')'

2- *Complete symbols, formulas, terms, closed formulas (sentences), and closed terms*

2.1- a) The variables and constants with a superscript r are complete symbols of type r. The occurrence of a variable is real in the complete symbol it constitutes. b) If z is a n-ary constant, and z_1, ..., z_n variables of type ι or constants (of whatever arity), then

$$\ulcorner z\, z_1 \ldots z_n \urcorner$$

is a complete symbol of type π_0. The occurrences of variables are real in it.

2.2- If c_1 and c_2 are complete symbols of types π_n, $\pi_{n'}$ respectively, then

$$\ulcorner(\neg\, c_1)\urcorner \text{ and } \ulcorner(c_1 \Rightarrow c_2)\urcorner$$

are complete symbols of types π_n, $\pi_{n''}$ respectively, where $n'' = \max(n, n')$. The occurrences of variables which were real (resp. apparent) in c_1 or c_2 remain so in the new complete symbols.

2.3- If c is a complete symbol of type π_n, and v a variable of type r having a real occurrence in c, then

$$\ulcorner(\forall v)c\urcorner$$

is a complete symbol of type $\pi_{n'}$, where $n' = \max(n, 1 + \mathrm{od}(r))$. The occurrences of variables which were real (resp. apparent) in c remain so in the new complete symbol, except for the occurrences of v which were real, which are now apparent.

2.4- If c is a complete symbol of type π_n, and v_1, ..., v_k variables of types r_1, ..., r_k respectively having real occurrences in c, then

$$\ulcorner(\lambda v_1 \ldots v_k)c\urcorner$$

is a complete symbol of type $\langle n', r_1, \ldots, r_k\rangle$, where $n' = \max(n, 1 + \max_{1 \leq i \leq k} \mathrm{od}(r_i))$. The occurrences of variables which were real (resp. apparent) in c remain so in the new complete symbol, except for the occurrences of v_1, ..., v_k which were real, which are now apparent.

The Russellian hierarchy must, then, be viewed as being the faithful ontological counterpart of the linguistic hierarchy of closed terms and sentences of L*(**TTR**), or rather of the hierarchy obtained by considering as identical the alphabetic variants of the same complete closed symbol of this linguistic hierarchy. Limiting the constants of L*(**TTR**) to type ι (rule 2.1-a) corresponds to the constructive character of the Russellian hierarchy; the rule 2.1-b corresponds to the propositions of order 0 obtained from individuals through the operation of predication alone (nothing corresponds to this operation in the language L(**TTR**), which does not carry out the analysis so far).

5. The Axiom of Reducibility

Adapting otherwise well-known procedures to the particular features of the language L(**TTR**), one could define the symbols for disjunction, conjunction, (material) equivalence, existential quantification, identity, description, and one could allow oneself to engage in various different abuses of language, as concerns punctuation, the abstraction operator (to whose use one might prefer, when no confusion is possible, using the circumflex accent *à la* Russell), type indices (the elimination of which would give way to 'systematically ambiguous' symbols), the marking out of the scopes of descriptions, etc. One could, in that way, recapture the style of *Principia*. I will myself risk adopting such a style.

I will examine the definition of identity here, an opportunity for the axiom of reducibility to make its first appearance.

One might think of defining identity by stipulating Leibniz's law for each type:

$$(1) \qquad x = y =_{\mathrm{Df}} (\forall f)(fx \Leftrightarrow fy).$$

The problem is that quantification cannot range over all the functions of one argument for which x is a possible argument, but only over

2.5- If $c, c_1, ..., c_k$ are complete symbols of types $\langle n, r_1, ..., r_k\rangle, r_1, ..., r_k$, then

$$\ulcorner c\, c_1 \dots c_k \urcorner$$

is a complete symbol of type π_n. The occurrences of variables which were real (resp. apparent) in c or in one of the c_i remain so in the new complete symbol.

2.6- A complete symbol is a formula or a term depending on wether its type is or is not of the form π_n.

2.7- A complete symbol without any real occurrence of a variable is closed; a closed formula is a sentence.

those of a specific order, *for example* (this is just an example) those which are of the lowest order compatible with the type of x, and which Russell unfelicitously calls 'predicative'. Whence the Russellian definition [cf. 1908, 245; 1910, 57 and 168-169]:

$$(2) \qquad x = y =_{\mathrm{Df}} (\forall f)(f\,!x \Leftrightarrow f\,!y\,),$$

where 'x' and 'y' stand for complete symbols of type $r \in \mathrm{T_r}$, 'f' for a variable of type $\langle n, r\rangle$, and where the exclamation mark indicates the predicativity of f: $n = 1 + \mathrm{od}(r)$.

One surmises that such a definition will not yield the expected results unless quantifying over the predicative functions of x in some way comes to the same thing as quantifying over all the functions of x, whatever they may be. To make sure of this, one need only let any function of x be coextensive with some predicative function of x. For then, if two entities have the same predicative properties, they have the same properties plain and simple, they are indiscernible, and can be thought of as identical. The 'axiom of reducibility' will stipulate, in a general way, that any function is coextensive with some predicative function [cf. 1908, 242-243; 1910, 56-58 and 166-16]:

$$(3) \qquad (\forall f)(\exists g)(\forall x_1 \ldots x_k)(f\,x_1 \ldots x_k \Leftrightarrow g\,!x_1 \ldots x_k),$$

where 'x_1', ..., 'x_k' stand for variables of type r_1, ..., r_k respectively, 'f' for a variable of type $r = \langle n, r_1, \ldots, r_k\rangle$ with $n > 1 + \max_{1\le i\le k} \mathrm{od}(r_i)$), '$g$' for a variable of type $\langle m, r_1, \ldots, r_k\rangle$, and where the exclamation mark indicates the predicativity of g: $m = 1 + \max_{1\le i\le k} \mathrm{od}(r_i)$).

Independently of the definition of identity, the axiom of reducibility's support is still needed for the definition of classes of entities, that of binary (resp. ternary, etc.) relations between entities, that of classes of classes of entities, etc., etc. These definitions are 'contextual': they simply consist in indicating how the formulas containing the new symbols — of classes of entities, for example — can be paraphrased into formulas no longer containing them. The classes of entities are not themselves entities; they do not figure in the ontological inventory of the theory of types; they are but 'logical fictions', or, as Russell also says, 'logical constructions' (which one will take care not to confuse with the 'real constructions', so to speak, discussed above). Binary (resp. ternary, etc.) relations between entities, classes of classes of entities, etc., etc. have the same status. I will not bring up Russell's definitions here[7]; I will just note that, as in the case of identity, the 'objects' so defined only have the expected properties owing to the axiom of reducibility.

Individuals, classes of individuals, binary (resp. ternary, etc.) relations between individuals, classes of classes of individuals, etc., etc. constitute what Russell calls the 'extensional hierarchy'. This hierarchy does not have the complexity, the 'ramified' character, as one says, of the intensional hierarchy (that of individuals, propositions and functions). The extensional hierarchy is a 'simple' hierarchy. Its elements are the very ones one provides oneself with from the outset in the ordinary extensional simple theory of types, hierarchized in identical fashion, but obtained here in a very roundabout way, and assigned the ontological status of fictional entities.

6. On the Classical Criticisms of the Axiom of Reducibility

By appealing to the axiom of reducibility, one can work up and down the extensional hierarchy and everything actually proceeds as in the extensional simple theory of types. And, as has been well-known since at least the end of the twenties, as long as one remains within the extensional simple theory of types, one is safe: all of classical mathematics can be reconstructed there. Indeed, to do this, one needs an axiom one would have liked to do without, the axiom of infinity (not to mention the axiom of choice), but, once one has it, things work out. However, the possibility of a similar reconstruction within the Russellian theory is in turn only opened up through the prior positing of another even more controversial axiom, the axiom of reducibility.

Much has been said about Russell's axiom of reducibility, and sometimes, it was Russell who was the first to speak up.

Poincaré [1909] confessed that he did not really understand what it was about.

Among those who thought they understood, some questioned the nature of the axiom of reducibility. It was no more of a logical axiom than was the axiom of infinity. Wittgenstein [1913b] was, no doubt, the first to attempt to demonstrate it, cryptically, in a letter to Russell at the end of 1913. But, be Wittengstein right or wrong, that did not preclude its being true in our world, and the ability to use it correspondingly.

---

7 For these definitions (revised), cf. [Rouilhan 1996, chap. VI, §§ 19-22].

The axiom of reducibility has also been criticized for being an *ad hoc* axiom, adopted, not because of any particular self-evidence, but for the sake of the cause. But Russell was well aware of this, and this provided him with an opportunity to reconsider the question of the epistemological grounds for choosing axioms [cf. Russell 1907; 1910, 59-60): logic and mathematics were no less *inductive*, to his way of thinking, than were the natural sciences, whose principles were not worthy of consideration because of any particular self-evidence, but rather because of the expected nature of results that could not be obtained without them.

The axiom of reducibility has also, following Chwistek's lead [1921], been accused of making the Russellian theory of types powerless as regards the semantical paradoxes it was supposed to solve [cf. Chwistek 1922; Copi 1950]. But this accusation turned out to be unfounded [cf. Ramsey 1925; Chwistek 1929; Copi 1970; Church 1976; Myhill 1979].

Others, following Ramsey's lead [1925], have proclaimed the uselessness of an axiom designed to loosen the grip of ramification and, first of all, of the vicious circle principle, a principle itself useless and even false. This point deserves further consideration.

The logicians thus opposed to the vicious circle principle (notably Gödel [1944, *in* 1986-??, vol. II, 127-128] and [Quine, 1963, 2d ed., 242-243]) have argued that such a principle would only be valid relative to a universe of objects "constructed by ourselves". Their opposition to the principle has gone hand in hand with the idea that mathematical objects "exist independently of our constructions" just as physical objects do. Conversely, a friend of the vicious circle principle like Feferman [e.g. 1964] has always based his defense of the principle on the idea that abstract objects "do not exist outside of us" and are but "mental constructs". However, in Russell's theory of types, as Quine has noted, the [propositions and] functions are the values of primitive variables and are therefore not "[mental] constructs", but "entities which are there from the start". Therefore, Russell was wrong to apply the vicious circle principle to them.

At the heart of this criticism is the idea, thereafter accepted, that the vicious circle principle would justifiably be party to constructivism in the anti-realist, to be more specific the mentalistic sense, in which it is a question of it in this matter. It is clear that Russell did not think that at all, and he was quite right in that. Even though he hardly explained it, the vicious circle principle for propositions and functions had nothing to do, to his way of thinking, with any possible character of mental construction of these entities, but rather, no doubt, with the extreme severity of the identity criteria these intensions were supposed to meet. For the idea that an intension satisfying such identity criteria might be the value of its own apparent variables turns out, upon refection, to be hardly less shocking than the idea that it might be one of its own constituents, or the idea that a class or a set might be one of its elements, or an element of one of its elements, etc.

This latter kind of possibility is known to be ruled out in both the intended model of the extensional simple theory of types and the intended model of Zermelo's set theory; I will call the heuristic principle for this exclusion the 'principle of foundation', whose formal counterpart for set theory goes by the name 'axiom of foundation'. The concept of set in the sense intended by Zermelo's set theory is known by the name 'iterative concept of set'. I will use the term 'iterative concept of extension' to refer to both the concept of class or relation in the sense intended by the extensional simple theory of types and the concept of set in the sense intended by Zermelo's theory. Finally, since Poincaré [1905-06], any notion in keeping with the vicious circle principle is called 'predicative'. I will therefore say: *for sufficiently strict identity criteria, a predicative concept of intension is just as natural as an iterative concept of extension, and the vicious circle principle as the principle of foundation.*

7. Gödel's Criticism : on the Incompatibility of the Axiom of Reducibility with the Axiom of Infinity in any Constructive Interpretation of the Language of the Russellian Theory of Types

It remains for me to bring up a last criticism, which, if it is justified, is truly invalidating. This is Gödel's criticism.

There is something extremely disturbing about the idea that the axiom of reducibility and the axiom of infinity make it possible to reconstruct classical mathematics within the ramified theory of types, i.e. to reconstruct, within a logic supposed to be predicative (that is to say in keeping with the vicious circle principle), mathematics which, as we well know nowadays, is not! One suspects that there is something fundamentally wrong with the axiom of reducibility, that, together with the axiom of infinity, the axiom of reducibility reintroduces behind the scenes what ramification, and in the first place the vicious circle principle, have ruled out in principle, namely impredicativity. But how can this be with an axiom which obviously does not involve any type mistake?

But it is not in the direction of the language that one has to look. It is to its interpretation. As an intended interpretation, it must be constructive (in the realist sense); furthermore, in the case under consideration, it must satisfy the axiom of reducibility and the axiom of infinity. But can an interpretation satisfy both these axioms simultaneously while retaining this constructive character? That is the question. And Gödel's answer [1944] is *no*. (Naturally, it was not exactly to the question I am now asking that Gödel answered *no*. He did not formalize his idea, and he seems to have had in mind a theory slightly different from the one I have drawn from reading Russell)[8].

Here is what Gödel says:

> [The axiom of reducibility] in essence already mean[s] the existence in the data of the kind of objects to be constructed [...]. In the first edition of *Principia*, [...] the constructivistic attitude was, for the most part, abandoned, since the axiom of reducibility [...] together with the axiom of infinity makes it absolutely necessary that there exist primitive predicates of arbitrarily high types [...]. [T]he axiom of reducibility [...] (in the case of infinitely many individuals) is demonstrably false unless one assumes [...] the existence of classes or of infinitely many '*qualitates occultae*'. [Gödel 1944, 142, 143 and 152]

In the presence of the axiom of infinity, if Gödel is right — and I believe he is right — the axiom of reducibility is scarcely worth anything more, all things considered, than a contradiction. I do not know what proof Gödel had in mind. Perhaps the publication of the *Nachlass* will provide the answer. The *Collected Works* currently being published do not, for the time being, promise anything of the like.

Even worse, Charles Parsons, in his introduction to Gödel's 1944 article, thinks one can dismiss any such proof beforehand:

> Gödel's remarks about the axiom of reducibility show lack of sensitivity to the essentially intensional character of Russell's logic; the fact that every propositional function is *coextensive* with one of lowest order does not imply "the existence in the data of the kind of objects to be constructed", if the objects in question are concepts or propositional functions rather than classes . [...] Russell himself was closer to the mark in saying that the axiom accomplishes "what common sense effects by the admission of *classes*." [...] This insensitivity is quite common in commentators on Russell, but is somewhat surprising in Gödel [...]. [Parsons 1990, 112-113]

Be Gödel right or wrong, the proceedings Parson institutes against him for having missed the essentially intensional character of Russell's logic is absurd. Parsons seems to believe that Gödel thinks of the axiom of reducibillity as essentially affirming, for any function of one argument, for example, the existence of the corresponding

---

8 Cf. above, footnote 3, 2°.

class. He seems to believe that the 'objects to be constructed' Gödel speaks of can be but classes. He seems to forget Russell's *realist* constructivism, so to speak, regarding propositions and functions, and retain only Russell's *logical* constructivism (or fictionalism) regarding classes and relations, to which Gödel does not limit himself at all. Lack of comprehension is quite common among commentators on Gödel, but it is somewhat surprising in Parsons, even though it is true that Gödel himself failed *explicitly* to make the *radical* distinction between the two kinds of constructivism called for — unless it is primarily a matter of not understanding Russell himself[9].

I will say, to conclude, why I believe Gödel was right: by analogy with an otherwise better known situation. Let us forget about Russell. Let us take as our individuals the set of whole numbers N. Let us define M_0 as the class of arithmetical sets; and M_1 as the class of sets definable by a condition (without parameters) in which quantification is relative to N or to M_0. The question of knowing whether, in a constructive interpretation of L(**TTR**) satisfying the axiom of infinity, every function of order 2 of an individual argument, for example, is coextensive with a predicative function (in the sense involved in the axiom of reducibility) is analogous to the question wether $M_0 = M_1$. The answer is *no*. By transfinite induction, one can, indeed, define an ever growing series of classes M_α leading up, at a certain ω_1-th step, to the class Δ^1_1 of so-called 'hyperarithmetical' sets: $M_{\omega 1} = \Delta^1_1$. On the other hand, we know that there exist hyperarithmetical sets which are not arithmetical: $M_0 \neq \Delta^1_1$ (these results are due to Kleene [1955a; 1955b; 1959]). It follows from this, given the definition of M_α, that, already, $M_0 \neq M_1$.

It now remains to verify that this demonstration that $M_0 \neq M_1$ can actually be transposed into a demonstration for L(**TTR**) of Gödel's thesis. In this form, the thesis at least has the merit of being mathematically precise, and one has an idea of how to demonstrate it. But that does not mean, unfortunately, that what comes next is a routine affair.

If the thesis I am defending in the wake of Gödel is correct, it deals a death blow to the Russellian version of the ramified theory of types with the axiom of reducibility. Those wishing to rescue the

9 The above formulation of my criticism of Parsons corrects the one I gave in [1996, 274].

theory against Russell himself would have but to forgo the constructive character (in the realist sense) the latter required of the hierarchy of propositions and functions, forgo the intelligibility they had in the beginning, and admit '*qualitates occultae*', to use Gödel's expression, and even, more specifically, '*qualitates occultae*' *of arbitrarily high types*. Or better, if possible, contest the 'occult' character of these qualities[10].

References

Aczel, Peter

1987 *Non-well Founded Sets*, CSLI Lectures Notes n° 14.

Church, Alonzo

1976 Comparison of Russell's Resolution of the Semantical Antinomies with that of Tarski, *Journal of Symbolic Logic* 41, 747-60.

Chwistek, Leon

1921 Antynomie logiki formalnej, *Przeglad Filozoficzny* 24; 164-171 (english trans., "Antinomies of Formal Logic", *in* [McCall 1967]).

1922 Ueber die Antinomien der Prinzipien der Mathematik, *Mathematische Zeitschrift* 14, 236-243.

1929 Neue Grundlagen der Logik und Mathematik, *Mathematische Zeitschrift* 30.

Copi, Irving M.

1950 The Inconsistency or Redundancy of Principia Mathematica, Philosophical and Phenomenological Research 11, 190-199.

1971 *The Theory of Logical Types*, London : Routledge and Kegan Paul.

Feferman, Solomon

1964 Systems of Predicative Analysis, *The Journal of Symbolic Logic* 29, 1-30.

Gödel, Kurt

1944 Russell's Mathematical Logic, *in* [Schilpp 1944, 123-153] (reprinted. *in* [Gödel 1986-??, vol. II, 119-141])

1986-?? *Collected Works* (ed. by S. Feferman *et alii*), New York et Oxford : Oxford University Press, vol. I, *Publications 1929-1936*, 1986; vol. II, *Publications 1938-1974*, 1990; vol. III, *Unpublished Essays and Lectures*, 1995; vol. IV, forthcoming.

---

10 This is what E. Zahar and (a bit perversely) Ch. Chihara invited me to do at Nancy. The best I could do was to decline.

Kleene, Stephen C.

1955a Arithmetical Predicates and Function Quantifiers, *Transactions of the American Mathematical Society* 79, 312-340.

1955b Hierarchies of Number Theoretic Predicates, *Bulletin of the American Mathematical Society* 61, 193-213.

1959 Quantification of Number-theoretic Functions, *Compositio Mathematica* 14, 23-40.

McCall, Storrs (ed.)

1967 *Polish Logic 1920-1939*, Oxford : Clarendon Press.

Mirimanoff, Dimitri

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

Myhill, John

1979 A Refutation of an Unjustified Attack on the Axiom of Reducibility, *in* [Roberts 1979].

Parsons, Charles

1990 Introductory note to *1944* , *in* [Gödel 1986-??, vol. II].

Poincaré, Henri

1905-06 Les mathématiques et la logique, *Revue de Métaphysique et de Morale* 13 (1905), 815-835; 14 (1906), 17-34, 294-317 (abridged and modified version *in* [Poincaré 1908]).

1908 *Science et Méthode*, Paris : Flammarion.

Quine, Willard Van Orman

1963 *Set Theory and its Logic*, Cambridge, Mass : Harvard University Press (2nd ed. 1969).

Ramsey, Frank Plumpton

1925 The Foundations of Mathematics, *Proceedings of the London Mathematical Society*, series 2, 25 (1926), 338-384 (paper read before the London Mathematical Society in 1925; reprinted *in* [Ramsey 1978]).

1978 *Foundations — Essays in Philosophy, Logic, Mathematics and Economics* (ed. by D. H. Mellor), London and Henley : Routledge and Kegan Paul.

Roberts, G. (ed.)

1979 *Bertrand Russell, Memorial Volume*, London : G. Allen and Unwin.

Rouilhan, Philippe de

1996 *Russell et le cercle des paradoxes*, Paris : P.U.F.

Russell, Bertrand

1907 The Regressive Method of Discovering the Premisses of Mathematics, published *in* [Russell 1973].

1908 Mathematical Logic as Based on the Theory of Types, *American Journal of Mathematics* 30, 222-262 (reprinted. *in* [Russell 1956]).

1910 Cf. [Whitehead and Russell 1910-27, vol. I, 1st ed].

1956 *Logic and Knowledge — Essays 1901-1950* (ed. by R. C. Marsh), London : G. Allen and Unwin.

1973 *Essays in Analysis* (ed. by D. Lackey), London : G. Allen and Unwin.

Schütte, Kurt

1960 *Beweistheorie*, Berlin : Springer.

Wittgenstein, Ludwig

1913a Letter to Russell of September 5, 1913, *in* [Wittgentein 1961, 2nd ed., 123].

1913b Letter to Russell of the end of 1913, *in* [Wittgentein 1961, 2nd ed., 127-129].

1961 *Notebooks 1914-1916* (ed. by G. H. von Wright and G. E. M. Anscombe, with an english trans. by G. E. M. Anscombe), Oxford : B. Blackwell (2nd ed. 1979).

1977 *Culture and Value* (ed. by G. H. von Wright, trans. by P. Winch), Oxford : B. Blackwell.

Zermelo, Ernst

1908 Untersuchungen über die Grundlagen der Mengenlehre, I, *Mathematische Annalen* 59, 261-281.

1930 Ueber Grenzzahlen und Mengenbereiche — Neue Untersuchungen über die Grundlagen der Mengenlehre, *Fundamenta Mathematicae* 16, 29-47.