Synthese (2013) 190:525–547
DOI 10.1007/s11229-011-0047-x

Logic of paradoxes in classical set theories

Boris Čulina

Received: 15 December 2005 / Accepted: 9 November 2011 / Published online: 3 December 2011

Abstract According to Cantor (Mathematische Annalen 21:545–586, 1883; Cantor's letter to Dedekind, 1899) a set is any multitude which can be thought of as one ("jedes Viele, welches sich als Eines denken läßt") without contradiction—a consistent multitude. Other multitudes are inconsistent or paradoxical. Set theoretical paradoxes have common root—lack of understanding why some multitudes are not sets. Why some multitudes of objects of thought cannot themselves be objects of thought? Moreover, it is a logical truth that such multitudes do exist. However we do not understand this logical truth so well as we understand, for example, the logical truth $\forall x\, x = x$. In this paper we formulate a logical truth which we call the productivity principle. Rusell (Proc Lond Math Soc 4(2):29–53, 1906) was the first one to formulate this principle, but in a restricted form and with a different purpose. The principle explicates a logical mechanism that lies behind paradoxical multitudes, and is understandable as well as any simple logical truth. However, it does not explain the concept of set. It only sets logical bounds of the concept within the framework of the classical two valued $\in$-language. The principle behaves as a logical regulator of any theory we formulate to explain and describe sets. It provides tools to identify paradoxical classes inside the theory. We show how the known paradoxical classes follow from the productivity principle and how the principle gives us a uniform way to generate new paradoxical classes. In the case of *ZFC* set theory the productivity principle shows that the limitation of size principles are of a restrictive nature and that they do not explain which classes are sets. The productivity principle, as a logical regulator, can have a definite heuristic role in the development of a consistent set theory. We sketch such a theory—the cumulative cardinal theory of sets. The theory is based on the idea of cardinality

B. Čulina (✉)
Department of Mathematics, University of Applied Sciences Velika Gorica,
Zagrebačka Cesta 5, Velika Gorica, Croatia
e-mail: boris.culina@vvg.hr

of collecting objects into sets. Its development is guided by means of the productivity principle in such a way that its consistency seems plausible. Moreover, the theory inherits good properties from cardinal conception and from cumulative conception of sets. Because of the cardinality principle it can easily justify the replacement axiom, and because of the cumulative property it can easily justify the power set axiom and the union axiom. It would be possible to prove that the cumulative cardinal theory of sets is equivalent to the Morse–Kelley set theory. In this way we provide a natural and plausibly consistent axiomatization for the Morse–Kelley set theory.

Keywords Set theory · Paradoxes · Limitation of size principles

Logistic is not sterile;
it engenders antinomies.
H. Poincaré

1 The productivity principle

According to Cantor (Cantor 1883, 1899) a set is any multitude which can be thought of as one ("jedes Viele, welches sich als Eines denken läßt") without contradiction—a consistent multitude. Other multitudes are inconsistent or paradoxial. In this section we will formulate a simple logical criterion to distinguish between sets and proper or paradoxical classes. We have named it the productivity principle.

When we talk about some objects, it is natural to talk about collections of these objects as well. The same happens when we talk about collections themselves. But the moral of paradoxes is that we can not talk freely about their collections.

First of all, the language has to be made precise. Since we are talking about collections, we will use the language of first order logic $L = \{\in\}$ where $\in$ is a binary predicate symbol which has a clear intuitive meaning: we write "$x \in y$" to say that collection x belongs to the collection y. However, as it is well known, it is impossible to have all the collections in the domain of the language, but only some (the goal is to have as much as possible), which will be called **sets**. So the model of the language L will be the (intended) universe of sets V equipped with the membership relation $\in$. A priori, it is an arbitrary model of L. Each formula $\varphi(x)$ of the language L is assigned a collection of objects from V satisfying the formula. The collection will be denoted $\{x \mid \varphi(x)\}$:

$$a \in \{x \mid \varphi(x)\} \leftrightarrow V \models \varphi(a)$$

Such collections will be called **classes**. A "serious" universe should represent every such class by its object, set s with the property

$$a \in \{x \mid \varphi(x)\} \leftrightarrow V \models a \in s$$

or expressed in the language L itself:

$$V \models \forall x \, (x \in s \leftrightarrow \varphi(x))$$

But *no matter how we imagine the universe, there will be classes of its objects which cannot be represented by its objects*. An example is **Russell's class** $R = \{x \mid x \notin x\}$. Namely, it is a *logical truth* of the language L that there is no set R with property

$$\forall x \quad x \in R \leftrightarrow x \notin x$$

Indeed, let us suppose that R is a set. Then, investigating whether it is an element of itself, we get a contradiction:

$$R \in R \leftrightarrow R \notin R$$

Not only is the result intuitively unexpected but it is a *logical truth* which we certainly do not understand as well as, for example, the logical truth $\forall x\, x = x$. The goal of this article is to understand better the logic which does not permit some classes to be sets.

The description of a situation follows. To every set s, as the object of the language L, we can associate class $\{x \mid x \in s\}$. But, the opposite is not true. There are classes that can be described in the metalanguage, like Russell's class, which are not objects of the language L, therefore are not sets. We will call them **proper classes** or, more in accordance with basic intuition, **paradoxical classes**.

To simplify, we will unite the reasoning about sets and classes into language $LL = \{\in\}$. This language has the same vocabulary as L but the intended interpretation is different; its objects are classes of sets. The classes that can be represented with sets will be identified with sets. Other classes, which cannot be the objects of the language L, we will call *paradoxical or proper classes*. Formally, we define in LL

A is a **proper** or **paradoxical class** $\leftrightarrow \forall X(A \notin X)$
A is a **set** $\leftrightarrow$ A is not a proper class.

In order to make the translation from the original language L to the extended language LL easier, we will use capital letters for variables over classes and small letters for variables over sets. So, for example, the formula of language L "$\forall x \varphi(x)$" can be considered as an abbreviation of the formula "$\forall X(X\ is\ a\ set \rightarrow \varphi(X))$" of the language LL.

We consider classes as collections of sets. Every condition on sets determines the class of all sets satisfying the condition. If classes have the same members, we consider them equal. These ideas will be formulated in the following axioms in the language LL:

axiom of extensionality:

$$A = B \quad \leftrightarrow \quad \forall x(x \in A \leftrightarrow x \in B)$$

axiom schema of impredicative comprehension:

$$\exists A \forall x(x \in A \leftrightarrow \varphi(x))$$

where $\varphi(x)$ is a formula of the language LL which does not contain A as a free variable.

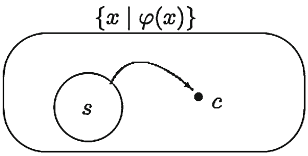
The consequence of the axioms is that each formula $\varphi(x)$ can be assigned a unique class $\{x \mid \varphi(x)\}$ of all sets x for which $\varphi(x)$ holds.

Let's note that to describe classes we will use impredicative comprehension, where $\varphi(x)$ is any formula of LL, instead of predicative comprehension where $\varphi(x)$ contains only quantification over sets. The impredicative comprehension gives us all classes we need regardless of the way in which we determine them (for details see Fraenkel et al. 1973, p. 119). To facilitate means of expression, we will take a maximalist approach to classes. The basic problem of any set theory is not what classes but rather what sets there are, and our position about classes is irrelevant for this problem. Almost all the reasoning (more precisely, all the reasoning which needs only predicative comprehension) can be translated either in the language L or as the reasoning about the language L, so we do not need to mention classes in any way. Such approach would minimize assumptions but it would complicate means of expression. Because of this we chose the language LL with the described axioms. The translation to L will be done only in some essential situations where it facilitates better understanding of the obtained results. Technique of the translation is well known—a discourse about the class $C = \{x \mid \varphi(x)\}$ can be understood as an abbreviation of a discourse about the formula $\varphi(x)$. We will need the following translations:

$s \in \{x \mid \varphi(x)\} \mapsto \varphi(s)$
$\{x \mid \varphi(x)\}$ is a set $\mapsto \exists s \forall x (x \in s \leftrightarrow \varphi(x))$
$\{x \mid \varphi(x)\}$ is a paradoxical class $\mapsto \neg\exists s \forall x (x \in s \leftrightarrow \varphi(x))$

With this terminology we only modeled a necessity for differentiation between paradoxical classes and sets. Now we will formulate a simple logical criterion to distinguish between sets and paradoxical classes.

The condition that there is no set of all objects that satisfy formula $\varphi(x)$ $(\neg\exists s \forall x (x \in s \leftrightarrow \varphi(x)))$ can be expressed in a logically equivalent way which shows an elementary logical mechanism that does not permit such a set. This logical biconditional we will call **the productivity principle**. According to that principle *paradoxical classes are exactly those classes for which there is a productive choice i.e a way to choose for every subset s of the class an object c of the class which is out of s*:



Proposition 1 *For $\varphi(x)$ a formula of L the following is a logical truth of L:*

$$\neg\exists s \forall x (x \in s \leftrightarrow \varphi(x)) \leftrightarrow (\forall s (\forall x (x \in s \rightarrow \varphi(x)) \rightarrow \exists x (x \notin s \text{ and } \varphi(x))))$$

Proof

($\rightarrow$): Let s be such that $\forall x (x \in s \rightarrow \varphi(x))$. According to the assumption on the left side of the biconditional there is an x such that $x \in s \leftrightarrow \varphi(x)$ isn't true. From these two conditions it follows that $\varphi(x) \rightarrow x \in s$ is not true. So $\varphi(x)$ is true and $x \notin s$.

($\leftarrow$): Assume the right side of the biconditional. If there is an s such that $\forall x(x \in s \leftrightarrow \varphi(x))$ $(*)$ then $\forall x(x \in s \rightarrow \varphi(x))$. Using the right side of the biconditional, we can conclude that there is an x such that $x \notin s$ and $\varphi(x)$. However, $(*)$ yields a contradiction—$x \notin s$ and $x \in s$. □

In the language LL this principle has more simple formulation but it is not a logical truth anymore. Namely, the concept of classes as collections of basic objects puts a minimal condition (extensionality of classes) on the principle.

Proposition 2 *With the assumption of the axiom of extensionality:*

C is a paradoxical class $\leftrightarrow$ for every set $s \subseteq C$ there is $x \in C \setminus s$.

Proof Suppose that C is a paradoxical class. Then for every set $s \subseteq C$ s is not equal to C (because C is not a set). Hence, using the axiom of extensionality, we can conclude that s is a proper subset of C i.e. there is $x \in C \setminus s$. Conversely, assume the left side of the biconditional. Suppose that C is a set. Then we can take C for a subset s of C. Therefore there is $x \in C \setminus C = \emptyset$, a contradiction. Hence, C is a paradoxical class. □

A condition for C that for every set $s \subseteq C$ there is $x \in C \setminus s$ has been formulated for the first time by Russell (Russell 1906) in his generalized contradiction. While Russell's formulation demands the existence of a definable operation which for a given subset of a class gives a new element of the class (a productive operation on the class), for the productivity principle the only thing that matters is the existence of a new object (a productive choice on the class). Moreover, Russell's view on the meaning of the principle is different (see 3 Limitation of size principles).

The productivity principle is a logical truth scheme in the language L. This means that we assume nothing about sets and that all of our assumptions about metalanguage LL of classes, introduced to facilitate the discussion, are irrelevant for the principle. The principle tells nothing specific about sets but it puts logical bounds on every theory of sets. *The productivity principle is also a simple logical truth*, because we can understand it almost as well as the logical truth $\forall x\, x = x$. Moreover it gives a simple logical mechanism of productive choice which prevents classes with such mechanism to be sets. A particular theory of sets postulates what sets there are and what operations over sets there are. The productivity principle logically translates this information into information about what paradoxical classes there are—they are collections on which the postulated (by means of the theory) fund of sets and operations enables a productive choice. We will show in the next section how to produce, using this principle, the known paradoxical classes and find new paradoxical classes. Basically, the principle says that we cannot have all imaginable sets and all imaginable operations at the same time. Then we could imagine an operation which when applied to every set gives a new element and we could also imagine a set that is closed under the operation. This is a contradiction in itself in the same way as the classical puzzle of the omnipotence of God. The basic religious intuition is that God is omnipotent, but then he could make a stone which he could not move. The omnipotence requires that he is able to move any stone and that he is able to make an unmovable stone. And this is a contradiction. The same conflict of basic intuition and logical bounds appears in the naive set theory.

In the sequel we will show how the productivity principle produces paradoxical classes, how it helps analyze the limitation of size principles and how we can use it as a guide in the construction of a theory about sets.

2 Finding paradoxical classes

In this section we will show how the known paradoxical classes follow from the productivity principle and how the principle gives us a uniform way to generate new paradoxical classes—*we must look for classes having a productive choice on themselves*. Each paradox (a discovery of paradoxical class) will be presented in two ways—in a uniform way, by establishing a productive choice on the class (which makes it paradoxical by means of the principle), and in a direct way, by means of repeating the proof of the productivity principle (to assume that the class is a set and to get a contradiction by applying a productive choice to the class). The first way displays a system to find paradoxical classes and explicates the logic behind them. The second way is more direct and explicates a paradoxicality of such classes with regard to primary intuition.

On the logical basis there is only one operation—identity, $s \subseteq V \mapsto s \in V$, so logical paradoxes are associated with classes where identity is productive:

$$s \subseteq C \mapsto s \in C \setminus s$$

Russell's class is such a class.

Russell's class $R = \{x \mid x \notin x\}$.

Uniform way. Let $s \subseteq R$. If $s \in s$ then $s \in R$, so $s \notin s$, a contradiction. Therefore $s \notin s$. However, then $s \in R$. So $s \in R \setminus s$ i.e. identity is productive on R.

Direct way. Suppose that R is a set. If $R \in R$ then $R \notin R$. So, $R \notin R$. But from the specification of R it follows that $R \in R$, a contradiction.

The class of not-n-cyclic sets, the class NI of sets which are not isomorphic to one of its elements, and the class of grounded sets are all classes on which identity is productive. Proofs are given in Appendix. We will establish here some relationships between such classes.

Let's note that the universe, in absence of other postulates, is not a class on which identity is productive, because we can imagine sets which contain themselves as members, for example $\Omega = \{\Omega\} \in \Omega$.

If we write down the productivity condition of identity on a class C in a more set-theoretical terminology we have

$$P(C) \subseteq C \cap R$$

where R is Russell's class. From that it is easy to get the following results:

Proposition 3 *1. The intersection of classes on which identity is productive is a class on which identity is productive.*

2. *If we assume the axiom of dependent choices, the class of ungrounded sets WF (for definition see Appendix) is the least class on which identity is productive.*
3. *Every transitive class on which identity is productive is a subclass of Russell's class.*

Proof 1. Let $P(C_i) \subseteq C_i \cap R, i = 1, 2$. Then $P(C_1 \cap C_2) \subseteq P(C_i) \subseteq C_i \cap R$, so $P(C_1 \cap C_2) \subseteq (C_1 \cap C_2) \cap R$.

2. From the groundedness of elements of WF, and using the axiom of dependent choices we can infer the induction principle: if the formula $\varphi(x)$ has the inductive property $\forall x \in y\varphi(x) \rightarrow \varphi(y)$ then $\forall x \in WF\varphi(x)$. Using this principle it is easy to prove that every class C on which identity is productive contains WF. Indeed, let $\forall x \in y\, x \in C$. Then $y \subseteq C$, so $y \in P(C) \subseteq C$, therefore $y \in C$. According to the induction principle we can conclude that every x from WF is in C.
3. If C is transitive then $C \subseteq P(C) \subseteq C \cap R \subseteq R$. □

However, in spite of these results we generally recognize such classes through their intensional characteristics. All classical examples we have mentioned are like that.

Other known paradoxes have the origin in the productivity of other operations. Their list is as follows:

Class *ORD* of all ordinals.

Uniform way. The productive operation on Ord is to get the first ordinal greater of all the ordinals from a given set. Namely, from the theory of ordinals it follows that for every set s of ordinals such ordinal exists, let's name it s^+. So, $s^+ \in ORD \setminus s$.

Direct way. Suppose *ORD* is a set. Then there is the first ordinal out of *ORD*, However, it is impossible since, by definition of *ORD*, *ORD* contains all the ordinals.

Class *CARD* of all cardinals.

Uniform way. From the theory of cardinals it follows that for every set of cardinals s there is the first cardinal s^+ greater of all the cardinals from s, so $s^+ \in CARD \setminus s$.

Direct way. Suppose that *CARD* is a set Then there is the first cardinal $CARD^+$ out of *CARD*. However, this is impossible since *CARD* contains all cardinals.

The universe is paradoxical if we assume the subset axiom. Usually, paradoxicality is proved by means of reduction to paradoxicality of Russell's class R. But we can prove it using a suitable productive operation.

Uniform way. We will show that diagonalization $s \mapsto \Delta s = \{x \in s \mid x \notin x\}$ is a productive operation on the universe (it is enabled by the subset axiom). Indeed, a condition on Δs of being an element of itself is

$$\Delta s \in \Delta s \quad \leftrightarrow \quad \Delta s \in s \quad \text{and} \quad \Delta s \notin \Delta s$$

If $\Delta s \in s$ then we have a contradiction:

$$\Delta s \in \Delta s \leftrightarrow \Delta s \notin \Delta s$$

Therefore $\Delta s \in V \setminus s$.

Direct way. Suppose that the universe V is a set. Then ΔV is a set out of V, which is impossible because V contains all sets. Let's note that $\Delta V = R$ so it is common to obtain a contradiction by showing that R is not a set.

We can find another productive choice on the classical universe of sets. The partitive set operation enables the choice. By Cantor's theorem there are more sets in $P(s)$ than there are in s, so there is an x in $P(s)$ out of s. Therefore we have a productive choice of $x \in V \setminus s$. If we transform this argument into a direct proof we obtain another standard proof for paradoxicality of V. Namely, if V is a set then it contain $P(V)$ as a subset, but it is impossible because $P(V)$ contains by means of Cantor's theorem more elements than V.

In a narrow connection with diagonalization and Cantor's theorem is the Šikić's class (Šikić 1986). Its paradoxicality is proved in Appendix.

In the same way for a given fund of operations the productivity principle gives instructions for generating *new* paradoxical classes—*we need to look for classes on which operations enable productive choice*.

class NWF **of all ungrounded sets**.

If we assume that there is an ungrounded set, that the universe is a paradoxical class, and the axioms of union and pair, then we can show that this class is paradoxical.

Uniform way. Let $s \subseteq NWF$. If $s = \emptyset$ then we get an element of NWF (such an element exists by means of the assumption of existence of ungrounded sets) for a new element of the class. If $s \neq \emptyset$ we get $x_1 \in s$ and $x_2 \notin \cup s$ (such an element exists because the universe is not a set). We claim that $\{x_1, x_2\}$ is a productive choice on NWF. Really, if $\{x_1, x_2\} \in s$ then $x_2 \in \cup s$, contrary to the choice of x_2. So $\{x_1, x_2\} \notin s$. But because of ungroundedness of x_1 there is an ungroundedness of $\{x_1, x_2\} \ni x_1 \ni \ldots$. Therefore $\{x_1, x_2\} \in NWF \setminus s$.

Direct way. Let NWF be a set. Under the assumption $NWF \neq \emptyset$ there is $x_1 \in NWF$. Because $\cup\, NWF$ isn't equal to the universe there is $x_2 \notin \cup\, NWF$. If $\{x_1, x_2\}$ belongs to NWF then x_2 belongs to $\cup\, NWF$, contrary to the choice of x_2. Therefore $\{x_1, x_2\} \notin NWF$. But x_1 is ungrounded so there is an ungroundedness of $\{x_1, x_2\} \ni x_1 \ni \ldots$,that is to say $\{x_1, x_2\} \in NWF$, a contradiction.

The proof is valid under the weaker assumptions, too. It is enough to suppose, beside the axiom of pair, that there exists an ungrounded set and that for every set s of ungrounded sets $\cup s \neq V$.

For the given function $F : V \longrightarrow V$ how can we find a class on which it is productive? If we assume that F is injective, such class is the following

class of all values of injective function F **which do not belong to their argument** $I = \{F(x) \mid F(x) \notin x\}$.

Uniform way. We will show that F itself is productive on I. Let $s \subseteq I$. If $F(s) \in s$ then $F(s) \in I$. Using the specification of I this means that there is an x such that $F(s) = F(x) \notin x$. But the injectivity ensures that x is equal to s, so we get that $F(s) \notin s$, and this contradicts the assumption. Therefore, $F(s) \notin s$. However, it implies $F(s) \in I$. Hence, $F(s) \in I \setminus s$.

Direct way. Let I be a set. If $F(I)$ belongs to I, then using the specification of I it means that there is an x such that $F(I) = F(x) \notin x$. But using injectivity of F x is equal to I, so we get $F(I) \notin I$.Therefore, $F(I)$ doesn't belong to I. But then, using the specification of I, $F(I)$ belongs to I; a contradiction.

Here are some examples of such functions and associated paradoxical classes:

1. $x \mapsto \{x\}$: $I = \{\{x\} \mid \{x\} \notin x\}$ (the existence of a singleton is assumed)

2. $x \mapsto \{x, a\}$: $I = \{\{x, a\} \mid \{x, a\} \notin x\}$ (The existence of a pair is assumed)
3. $x \mapsto x \cup \{x\}$: $I = \{x \cup \{x\} \mid x \cup \{x\} \notin x\}$ (The existence of singleton and union is assumed, as well as the foundation axiom)
4. $x \mapsto (x, a)$: $I = \{(x, a) \mid (x, a) \notin x\}$ (the existence of a pair is assumed)
5. $x \mapsto (x, a, b \ldots)$: $I = \{(x, a, b, \ldots) \mid (x, a, b, \ldots) \notin x\}$ (the existence of a pair is assumed)
6. $x \mapsto P(x)$: $I = \{P(x) \mid P(x) \notin x\}$ (the existence of a partitive set is assumed)
7. $x \mapsto F[x]$, for injective F: $I = \{F[x] \mid F[x] \notin x\}$ (the replacement axiom is assumed)

....

Paradoxes connected with injective functions can be generalized by switching from functions to systems of rules (Aczel 1977) which are closer to the idea of a productive choice. The system of rules is a concept which connects different operations in a unity, heterogeneous ways of getting new elements into a unique system. Abstractly formulated, the **system of rules** is every class of ordered pairs. The intended interpretation is suggested by means of suitable terminology. If an ordered pair (s, x) belongs to the system of rules Φ we write $\Phi : s \vdash x$ and we say that in system Φ object x is derived from set of objects s. When it is clear from context what system Φ is considered, its label will be dropped. Examples of such systems are formal proof systems, or the system for generating natural numbers. We usually use them for generating some objects (let's say theorems or numbers) using an iteration of rules beginning from some initial objects (let's say axioms or number 0). We say that **the system is productive on** C when for all $s \subseteq C$ there is an x such that $s \vdash x \in C \setminus s$. Although we can generally infer more than one object from a given set (there is no uniqueness) the concept of injectivity can be easily generalized to systems—we say that a system is **deterministic** if every object can be derived from one set at the most:

$$s_1 \vdash x \quad \text{and} \quad s_2 \vdash x \ \rightarrow \ s_1 = s_2$$

We say that a system is **global** $\leftrightarrow$ $\forall s \exists x\, s \vdash x$.

Let's show that for a global deterministic system the following class is paradoxical:

class of all objects which do not belong to sets they are derived from in a global deterministic system Φ:

$$I = \{x \mid \exists a (a \vdash x \quad \text{and} \quad x \notin a)\}.$$

Uniform way. Let $s \subseteq I$. Because of globality there is an x such that $s \vdash x$. If x belongs to s then it belongs to I, so there is an a such that $a \vdash x$ *and* $x \notin a$. However, using determinism of the system, $a = s$, so $x \notin s$, and this is a contradiction. Therefore $x \notin s$. But then using the specification of I, $x \in I$. Therefore $x \in I \setminus s$.

Direct way. Let I be a set. Then there is an x such that $I \vdash x$. If x belongs to I then, using the specification of I and determinism of the system, x does not belong to I, a contradiction. Therefore x doesn't belong to I. But then, using the specification of I, it belongs to I, and this is a contradiction, too.

Thus, for example, by combining the previous injective operations we can get, with some caution, new deterministic systems and associated paradoxical classes. For example, combining operations $x \mapsto \{x\}$ and $x \mapsto P(x)$ we get a system which is deterministic because the situation $\{x\} = P(y)$ can be obtained only in one way, for $x = y = \emptyset$.

The concept of productivity can be generalized into the concept of a monotonic operator, too. For an operator ϕ we say that it is monotonic if for any set a and b $a \subseteq b \rightarrow \phi(a) \subseteq \phi(b)$. With each system Φ we can associate a monotonic operator which maps every set x to the set of all objects inferred from some subset of x.

$$s \mapsto \phi(s) = \{x \mid \exists a(a \subseteq s \quad \text{and} \quad a \vdash x)\}$$

Monotonic operator ϕ is productive on C if for every $s \subseteq C\, \phi(s) \cap C \setminus s \neq \emptyset$. Let's note that a system needs not to be productive for the associated operator to be productive. For example the standard system for generating ordinals is such a system:

$$\{\alpha\} \mapsto \alpha \cup \{\alpha\}$$
$$s \mapsto \cup s$$

Paradoxicality of *ORD* doesn't follow from the system because if α is the successor ordinal then $\cup\alpha \in \alpha$. But in α exists its predecessor β, so from $\{\beta\} \subseteq \alpha$ we can infer $\alpha \notin \alpha$. By generalization of the previous notice we can prove that the associated monotonic operator is productive and from this we can infer paradoxicality of *ORD* from it.

But is there any systematic way to find a class on which a given system of rules is productive, a class specified by some extensional means and not intensional? Some results in that direction follow.

Classes on which a given operation F is productive need to be looked for amongst classes closed under the operation. The first such class is $I(F)$ (so called the least fixed point of F), the class of objects given by iterative application of F beginning with the empty set. It can be described in various ways and each of them needs some postulates. We will describe it as the intersection of all classes closed under F:

$$I(F) = \cap\{C \mid \forall s \subseteq C \quad F(s) \in C\}$$

The description, although intuitively clear, can't be considered literally because elements of classes can't be classes, but as an abbreviation for

$$I(F) = \{x \mid \forall C((\forall s \subseteq C\; F(s) \in C) \rightarrow x \in C)\}$$

In the same way we will consider other similar descriptions which will be used soon. They are based on the impredicative principle of comprehension of classes which is discussed at the beginning of the article.

Equally, for system Φ we define

$$I(\Phi) = \cap\{C \mid \forall s \forall x(s \subseteq C \quad \text{and} \quad \Phi : s \vdash x \rightarrow x \in C)\}$$

The next proposition shows that the problem of productivity of a given operation F or a given system of rules Φ on some class is reduced to the problem of its productivity on its least fixed point:

Proposition 4 *(i) For function F there is a class on which it is productive $\leftrightarrow$ it is productive on $I(F)$.*
(ii) For system of rules Φ there is a class on which it is productive $\leftrightarrow$ it is productive on $I(\Phi)$

Proof (i): The direction $\leftarrow$ is trivial. Let's prove the direction $\rightarrow$. Let F be productive on C. Then C is closed under F, so $I(F) \subseteq C$. Let $s \subseteq I(F)$. Then $s \subseteq C$, so, because of productivity on C, $F(s) \notin s$. However, because $I(F)$ is closed under F, $F(s) \in I(F)$, so $F(s) \in I(F) \setminus s$. Therefore, F is productive on $I(F)$. For (ii) the proof is similar. □

Corollary 1 *1. If $F : V \longrightarrow V$ is injective then $I(F)$ is a paradoxical class.*
2. If Φ is a global deterministic system of rules then $I(\Phi)$ is a paradoxical class.

Proof We have shown that an injective function F is productive on $\{F(x) \mid F(x) \notin x\}$ and that a global deterministic system of rules Φ is productive on $\{x \mid \exists a(a \vdash x\ i\ x \notin a)\}$. Using the previous proposition it means that they are productive on their least fixed point. □

For example, the least fixed points of all previously mentioned injective operations and global deterministic systems are paradoxical classes.

3 Limitation of size principles

According to the productivity principle paradoxical classes are exactly those classes on which there is a productive choice. The existence of such a choice doesn't mean necessary that a theory about sets (or the underlying conception of set) assures enough sets and operations. It can even mean the opposite, that there are not enough sets. This is exactly the case with the limitation of size principles. We will see that *the limitation of size principles assure a productive choice of a certain kind on paradoxical classes because they are of a restrictive nature and don't allow more sets*.

The analysis of paradoxical classes in *already existing* Cantor's set theory (which mathematicians from the beginning of the century as well as today used and needed) extracted their common property—they are in a sense "too big". In Cantor it is in the sense of how many elements there are in a class (the cardinal sense), in Mirimanoff in the sense of how many stages there are in a class formation (the cumulative sense). The idea of "too big" is also presented in Zermelo and Fraenkel, although not in such a definite sense. However, all the variations are in a harmony with the phrase "too big" as it is used in ordinary language, so we will use the phrase in a general context. Of course, if *all known* paradoxical classes are "too big", it does not mean that *all* paradoxical classes are as such. Thus the observation is formulated as the **limitation of size hypothesis** (Hallett 1984, p. 176): *all paradoxical classes are (in some sense)* *too big*. Because the idea of "too big" suggests an explanation why some classes are not sets, it is taken as a criterion to distinguish between sets and paradoxical classes:

$$C \text{ is a paradoxical class } \leftrightarrow C \text{ is too big}$$

We will use the word "small" instead of "not too big", so in the contraposition form we have

$$C \text{ is a set } \leftrightarrow C \text{ is small}$$

For every meaning of "too big" we have a corresponding principle, called the **limitation of size principle**. As every such principle sounds as an explanation what a set is, it can be taken as a basis of a definite conception of set, which we will call here the **limitation of size conception**.[1] However, without other explanations of what is small (= not too big) the limitation of size principles do not carry any information about sets; we have only new names for the old concepts ("small" for "set" and "too big" for "paradoxical class"). Other explanations or postulates must declare what is small. In Cantor, "small" means enumerated by some ordinal, and in Mirimanoff "small" means to be a subset of some stage V_α of the cumulative hierarchy. Of course, what is small depends on what ordinals i.e. what stages there are. Taken together these other postulates give us the meaning of the limitation of size principle in one direction:

$$C \text{ is small } \rightarrow C \text{ is a set } (\text{postulates on what is small})$$

It means that the limitation of size principle carries specific information only in the opposite direction, so we will consider the limitation of size principles to be the following statements:

$$C \text{ is too big } \rightarrow C \text{ is a paradoxical class}$$

or in the contraposition form:

$$C \text{ is a set } \rightarrow C \text{ is small}$$

We will now analyze the meaning of the limitation of size principles formulated in this way. Other postulates say to us what sets there are. Acording to the productivity principle these sets and operations on them give us productive choices on some classes and make them paradoxical. Bertrand Russell was the first to notice this in his correspondence with Jourdain (see Hallett 1984, p. 180). Russell starts from an observation that all known paradoxical classes have a common property which he expressed in **Russell's generalized contradiction**:

[1] in the literature on set theory it is common to use this term in the specific meaning where "too big" is in Cantor's sense.

Let there be a function symbol F and a formula $\varphi(x)$ such that

$$\forall s[\forall x(x \in s \to \varphi(x)) \to F(s) \notin s \text{ and } \varphi(F(s))]$$

Then $\{x \mid \varphi(x)\}$ is a paradoxical class.
We can say it in a more readable way:

Let there be a function F and a class C such that for every $s \subseteq C$ $F(s) \in C \setminus s$. Then C is a paradoxical class.

This condition for paradoxicality is the same as in the productivity principle, except that in Russell there must be some definable operation that gives productive choices. Also, for Russell the condition was not a *criterion* for paradoxicality but a step in arguing limitation of size hypothesis. He observed that not only *ORD* has such a property, but that all other classes with such a property contain *ORD*, or better to say, its isomorphic copy. Namely, in a paradoxical class the sequence of different elements can be reproduced (it is guaranted by the postulates on what is small):

$$\emptyset,\ F(\emptyset),\ F(\{\emptyset, F(\emptyset)\}), \ldots$$

The next new element in the sequence is the result of applying F on the set of previous elements. So we get an ordered structure isomorphic to *ORD*. Even more, if we take Russell's class and identity we get just "official" von Neumann's ordinals. In that way Russell argued limitation of size hypothesis showing that all the known paradoxical classes contain an isomorphic copy of *ORD*, so they are too big likewise *ORD* is too big, and not small like sets postulated by the postulates on what is small. The limitation of size hypothesis saying that all paradoxical classes are too big means in the contraposition form that all small classes are sets. Thus, Russel's argument is the argument that makes plausible the consistency of the postulates on what is small enough to be a set.

Limitation of size principles go further. They say that *all classes that are not small are paradoxical classes i.e. there are no more sets except those that are small*. The postulates on what is small say what sets there are and limitation of size principles say that there are no more sets. Thus *limitation of size postulates are the closure postulates of the postulates on what is small*. Therefore, they are of a restrictive nature. On one side, according to the postulates on what is small, there are collections small enough to be sets. On the other side these postulates assure productive choices on some collections and make them paradoxical and too big. However, there are a lot of collections between these opposites which are not small in the sense of underlying conception on what small is and which ones are not paradoxical. Let's illustrate this on the cumulative conception. All collections which are built in the cumulative manner beginning with the empty set are sets. The collection which has only itself as a member, $\Omega = \{\Omega\}$ doesn't belong to the cumulative hierarchy but it is not too big in the intuitive sense. Furthermore it is not paradoxical. Empty set and Ω itself are the only subcollections of Ω. If we accept Ω as a set then Ω has two subsets. We can find an object in Ω outside $\emptyset$ (it is Ω) but not outside Ω. So, there is no productive choice on Ω. However, the limitation of size principle will proclaim Ω too big i.e. paradoxical. According to the productivity principle this means that there is a productive choice on Ω. However, this choice is trivial and enabled by the decision that Ω is not a set. Namely, because Ω is not a set the only subset is $\emptyset$ and now we have a productive choice. As another example let's take the whole universe V. Although it is not the element of the cumulative hierarchy it is not paradoxical in the absence of other postulates neither. The consequence of the postulates on what is small is that the whole cumulative hierarchy WF is paradoxical class, as it is well known (see Appendix). If we allow some collections which are outside the cumulative hierarchy, like Ω, to be sets, then V is not necessarily paradoxical, but if we don't allow more sets, and this is the meaning of the limitation of size principle, then $V = Wf$ is a paradoxical class.

On the previous examples we can see that the cumulative limitation of size principle enables a productive choice because it doesn't allow some collections to be sets. Now we will show that this is generally so. We will show that the limitation of size principle enables a productive choice of a certain kind on paradoxical classes. Basically, this choice is a choice of an object x outside V_α, and it is a productive choice because the limitation of size principle doesn't allow collections outside the cumulative hierarchy to be sets.

Let's suppose some postulates about cumulative hierarchy (the postulates on what is small). The exact formulation of these postulates doesn't matter here. It is enough that they say that every subcollection of some stage V_α is a set. Then, from the cumulative principle of the limitation of size ("too big" means now not to be a subset of a some stage V_α)

$$\forall\alpha\exists x\, x \in C \setminus V_\alpha \ \rightarrow\ C \text{ is a paradoxical class}$$

it follows the statement about the existence of a certain kind of a productive choice on a paradoxical class:

$$C \text{ is a paradoxical class } \leftrightarrow \forall s \subseteq C\, \exists\alpha\exists x\, s \subseteq V_\alpha \text{ and } x \in C \setminus V_\alpha\ (*)$$

Proof Let's suppose the limitation of size principle. The direction $\leftarrow$ of $(*)$ follows from logic. Namely, according to the productivity principle we must prove that there is a productive choice on C and the left side of $(*)$ provides such a choice. To prove the direction $\rightarrow$ let's suppose that C is a paradoxical class and that $s \subseteq C$. From the limitation of size principle we can conclude by contraposition that there exists α such that $s \subseteq V_\alpha$ (there are no other sets!). However, C is a paradoxical class, so, according to the postulates about cumulative hierarchy, it is not a subcollection of V_α. Thus there is $x \in C \setminus V_\alpha$. Therefore, we have proved the direction $\rightarrow$ of $(*)$.

The same analysis can be carried out for the Cantor's cardinal limitation of size principle. The principle proclaims paradoxical all collections that cannot be enumerated by an ordinal (= too big collections):

There is no ordinal α and function F such that $C = F[\alpha]$
$\rightarrow$ C is a paradoxical class

Again, we have postulates about what is small (postulates about ordinals, the replacement postulate, etc.). We will call them postulates about enumerated collections. These postulates assure productive choices on some collections which makes them paradoxical and not enumerated by ordinal (*ORD*, the universe V, etc.). However, what is with the collection V_ω of all hereditary finite sets, for example? If we don't postulate the existence of infinite sets then V_ω is not enumerated, but there is nothing which makes it paradoxical. For any finite subcollection (therefore a set) s we can find an element of V_ω outside s (because s is finite and V_ω is infinite), but for the whole V_ω we cannot. If we close the conception of hereditary finite sets by corresponding cardinal limitation of size principle which says that there are no other sets then V_ω is not a set. The finite subcollections of V_ω are the only subsets and now we have a productive choice. Such a choice is again the consequence of the decision not to allow more sets. Again, we can prove that from the cardinal limitation of size principle (from forbidding some collections to be sets) follows the existence of a productive choice of a certain kind on paradoxical classes:

$$C \text{ is a paradoxical class } \leftrightarrow \forall s \subseteq C\, \exists F \exists \alpha \quad F[\alpha] = s \text{ and } F(\alpha^+) \in C \setminus s$$

We can repeat the same analysis for Zermelo's approach to paradoxes. In his approach "small" means to be a subcollection of a set (the subset axiom) and the relative character of the meaning of "small" is explicit here. The corresponding limitation of size principle is

$$\textit{there is not } s \text{ such that } C \subseteq s \to C \text{ is a paradoxical class}$$

However, the principle is now the truth of logic, therefore it says nothing about sets. Namely, in a contraposition form we need to prove that if C is a set, then there is a set s such that $C \subseteq s$. And the assumption gives such a set—it is just C. The presence of a productive choice on a paradoxical class follows directly, too. Because such a class C goes beyond every set, it goes also beyond any of its subsets, so for $s \subseteq C$ there is $x \in C \setminus s$.

Using the productivity principle we can analyze Fraenkel's limitation of size explanations of the concept of set which are rather loose. Explaining the difference between paradoxical classes and sets Fraenkel (1927), Fraenkel (1928) says that paradoxical classes are of *unbounded extension* and also that they *involve too much*. For the first metaphor of an unlimited extension we consider it similar to Zermelo's metaphor which we have already analyzed. Concerning the second metaphor of "too big involvement", as the main argument Fraenkel mentions that sets, as opposite to paradoxical classes, can be extended (the idea is present already in Cantor) and he illustrates this with diagonalization (which requires the subset axiom):

$$s \mapsto \Delta s = \{x \in s \mid x \notin x\} \notin s$$

Conversely to sets, neither Russell's class nor the universe are extended under diagonalization. For example, for $s \subseteq R \quad \Delta s = s \notin s$, so $\Delta s \in R$. It is not clear from Fraenkel's exposition whether he considers the closure under diagonalization to be a distinguishing property, or the closure under any operation which extends sets. The first thought is wrong because for example *ORD* is a paradoxical class which is not closed under diagonalization (for $s \subseteq ORD$ which is not an ordinal $\Delta s = s$ doesn't belong to *ORD*). The second thought has the confirmation in Fraenkel's metaphor of the "wall" according to which, contrary to sets, to constitute Russell's class "elements have to be taken from outside every wall, no matter how inclusive" (Fraenkel 1927). This gives a distinguishing criterion which is just the productivity principle and Fraenkel's explanation of the concept of set is a logical truth. So, it says nothing specific about sets.

Von Neumann's limitation of size principle is a variation of the cardinal limitation of size principle. In this approach "too big" means equipotent to the universe V of all sets:

$$C \textit{ is equipotent to the universe } V \leftrightarrow C \textit{ is a paradoxical class}$$

The postulates on what is small will tell what collection are not equipotent to the universe and hence are sets. Because *ORD* is a paradoxical class, it is equipotent to the universe V, and as such, by the postulates on what is small, to every paradoxical class. It means that all paradoxical classes have the same kind of a productive choice as *ORD* has.

4 Heuristic of the productivity principle

The productivity principle can guide us in the construction of the consistent set theory in the following way. When some sets and operations on sets are postulated the principle tells us what classes are paradoxical. We cannot proclaim any such class to be a set because we would get a contradiction. However, we can proclaim some other classes, which are not paradoxical by means of the principle, to be sets. So we get a richer theory. Now we can repeat the process and get more sets, and so on. If we do that in some systematic manner we can get a pretty rich theory. Of course, we cannot prove that such a theory is consistent but the procedure would convince us that it is the case. We will sketch one such a theory, called here **cumulative cardinal theory of sets** which realizes the corresponding cumulative cardinal conception of sets. Because of the construction its consistency seems plausible. Moreover, it inherits good properties from cardinal conception and cumulative conception of sets and as such it easily justifies ZFC axioms. Because of the cardinality principle it justifies easily the replacement axiom, and because of the cumulative property it justifies easily the power set axiom and the union axiom.

Before the construction we need to repeat that from the productivity of a function F follows that $I(F)$ is a paradoxical class (see 2 Finding paradoxical classes). From the assumption that $I(F)$ is a set and applying F on $I(F)$ we get a contradiction. For the productivity of F it is necessary that F is defined on every subset of $I(F)$. But some functions are such that it is not always fulfilled—there are given conditions of their applicability. If $I(F)$ belongs to the condition, although the function isn't defined on every subset of $I(F)$, we will again get a contradiction and therefore show that $I(F)$ is a proper class. Contrary, if $I(F)$ doesn't satisfy the condition the argument fails. Moreover, it seems plausible just the opposite—that $I(F)$ is a set. Namely, $I(F)$ is a collection generated by iterative application of F, so if the operation itself doesn't give a contradiction it would be surprising that some other operation or system of rules does so.

We will collect objects into sets in an order of growing cumulative cardinality of collecting. The collecting of the previous kind produces the least fixed point which will be considered as an initial place for the collecting of the next kind which has a greater cardinality. The basic idea is the following. Beginning with the empty set we can get all hereditary finite sets by finite and combinatory means, so their existence is indisputable. For example, we can generate them with the following system of rules:

$\emptyset \vdash \emptyset$
(we accept the empty set)

$\{x, y\} \vdash x \cup \{y\}$
(by adding to set x one element y we get again a set)

All sets generated in this way are hereditary finite (they are not only finite, but their elements are finite too, and elements of their elements too, etc.) and they make the least fixed point C_1 of the operation of *finite* collecting of objects to sets. However, C_1 is not finite, so the operation isn't applicable to it. Using the heuristic of the productivity principle we accept C_1 as a set. However, accepting it as a set we accept a countable collecting of objects into sets, too:

$$X \subseteq V \vdash X \in V \qquad \text{when there is an } F \text{such that } F[C_1] = X$$

where $F[S] = \{F(x) \mid x \in S\}$, and it need not to be that $S \subseteq dom(F)$. If $\exists F\ F[Y] = X$ we will write $X \preccurlyeq Y$ (X is dominated by Y or X is Y-enumerated)

This type of argument that accepting a set with some cardinality means accepting all collecting with that cardinality we will call the **cardinality principle**.

Beginning with objects from C_1 and iterating countable collectings (it includes finite collectings too) we get all hereditary countable sets. They make the least fixed point C_2 of the collecting. C_2 contains C_1 as a subcollection. But because $C_1 \notin C_1$ and $C_1 \in C_2$, C_2 is a richer collection:

$$C_1 \subset C_2$$

The story repeats itself now. Let's show using a modification of *Cantor's argument* that a countable collecting isn't applicable on C_2. If it isn't so then there is the function F such that $F[C_1] = C_2$. Let's consider $D = \{x \in C_1 \mid x \notin F(x)\}$. Elements of D are sets from $C_1 \subseteq C_2$ and $D = G[C_1]$, where G is identity function on D, so D is a set from C_2. However, using the assumption on F, there is $d \in C_1$ such that $F(d) = D$. This leads to a contradiction—$d \in F(d) \leftrightarrow d \notin F(d)$. Therefore, a countable collecting isn't applicable on the class of all hereditary countable sets C_2.

Because a countable collecting isn't applicable on class C_2 of all hereditary countable sets, using the heuristic of the productivity principle we can accept it as a set. However, C_2 has a greater cardinality, so accepting it as a set, using the cardinality principle we accept all collectings of that cardinality and the process goes on.

In the same way we solve every next finite jump because the previous proof remains valid when C_1 is substituted by C_α, C_2 by C_{α^+} (where α^+ is the next ordinal following α), and the expression "countable" by "C_α**-countable**" (enumeration by C_α).

Accepting C_α for a set, using the cardinality principle we accept any C_α-collecting of *already collected* objects, too. By means of the previous considerations C_α-collecting is not applicable on its fixed point C_{α^+} and using the heuristic of the productivity principle we accept C_{α^+} as a set. By repeating this process we get an ascending hierarchy of sets:

$$\emptyset = C_0 \subset C_1 \subset C_2 \subset \cdots \subset C_\alpha \subset C_{\alpha^+} \cdots$$

A transfinite jump is obtained by putting together all of the already accepted C_β-collectings, over all ordinals β less than the limit ordinal α. Such a collecting will be called $< C_\alpha$**-collecting**:

$$X \subseteq V \vdash X \in V \qquad \text{when there is } \beta < \alpha, \text{ such that } X \text{ is } C_\beta\text{-enumerated}$$

The least fixed point of the collecting gives the next limit member C_α of the hierarchy. From the definition of limit C_α it follows that C_α contains all previous stages and that it is not identical to any of them (because when it contains some stage C_β it contains also the next stage which C_β doesn't contain). Let's show that C_α isn't $< C_\alpha$-enumerated. If it is so, then there is the function F such that $F[C_\beta] = C_\alpha$ for some $\beta < \alpha$. As we did for a finite jump, let's consider $D = \{x \in C_\beta \mid x \notin F(x)\}$. Elements of D belong to $C_\beta \subset C_\alpha$ and D is C_β-enumerated, so it belongs to C_α. Therefore, there is $d \in C_\beta$ such that $D = F(d)$. However, this entails the contradiction: $d \in F(d) \leftrightarrow d \notin F(d)$.

Because $< C_\alpha$-collecting isn't applicable on the limit C_α and C_α is its least fixed point, using the heuristic of the productivity principle we accept C_α as a set. So we can continue the sequence through transfinite jumps, too:

$$\emptyset = C_0 \subset C_1 \subset C_2 \subset \cdots \subset C_\alpha \subset C_{\alpha^+} \cdots \subset C_\omega \subset \cdots$$

Every stage of the hierarchy contains greater cardinalities than previous stages and it is a measure for yet greater cardinalities. The hierarchy spreads as much ordinals as we can imagine. But as the construction gives greater cardinalities, at the same time it gives ordinals for next jumps.

Let's note some properties of the stage C_α. Because C_α is the least fixed point of collectings with smaller cardinality there is a corresponding induction principle for proving properties of C_α and a corresponding recursion principle for defining functions on C_α. Furthermore, the following sentences are true:

$$\begin{aligned} &S \subseteq C_\alpha \ \rightarrow \ S \in C_{\alpha^+} \\ &S \in C_\alpha \ \rightarrow \ S \subseteq C_\alpha \\ &S \in C_{\alpha^+} \quad \leftrightarrow \quad S \subseteq C_{\alpha^+} \ \text{ and } \ S \preccurlyeq C_\alpha \end{aligned}$$

For limit ordinal α

$$S \in C_\alpha \quad \leftrightarrow \quad S \subseteq C_\alpha \ \text{ and } \ S \preccurlyeq C_\beta \text{ for some } \beta < \alpha$$

Can we go further, by putting together *all* C_α-collectings for the final jump:

$$X \subseteq V \vdash X \in V \qquad \text{when there is an } \alpha, \text{ such that } X \text{ is } C_\alpha\text{-enumerated?}$$

If we could get new sets in this way then, by the same argument as in previous jumps, the universe V of all sets would be just a new stage for new jumps. This is impossible, by the very idea of the universe of all sets. Therefore the hierarchy of C_α gives all sets and it is closed under C_α-collectings:

$$V = \bigcup_{\alpha \in ORD} C_\alpha = \{S | \exists C_\beta (S \preccurlyeq C_\beta)\}$$

It means that

$$S \text{ is a set } \leftrightarrow \exists C_\alpha (S \preccurlyeq C_\alpha) \leftrightarrow \exists C_\beta (S \in C_\beta) \leftrightarrow \exists C_\gamma (S \subseteq C_\gamma)$$

This gives the limitation of size principle for the cumulative cardinal conception of set:

If C is not a subcollection of some stage C_α (=not dominated by some stage C_β) then C is a paradoxical class

Of course, this limitation of size principle is a closuring principle for this conception of set as well as other limitation of size principles are. For example, it excludes non-well-founded sets.[2]

We have described the **cumulative-cardinal conception of set** informally. Now, we will do it more formally. We will formulate axioms in the language LL (see 1 The productivity principle) and call them the **CC axioms**. For classes we accept the **axiom of extensionality** and the **axiom schema of impredicative comprehension**. Also, we accept the **axiom of choice**:

For every set $s = \{x_i | i \in a\}$ of nonempty disjoint sets there is a set c, called a choice set for s, such that for every $i \in a$ $x_i \cap c$ contains exactly one element.

The cumulative cardinal conception is neutral about the axiom of choice as well as cardinal and cumulative conceptions, because the axiom of choice is of a different nature. However, the cumulative cardinal conception permits the axiom of choice. By the axioms of union and subset (which are valid in the cumulative cardinal conception, as we shall see), if the collection c exists, then $c \subseteq \cup\{x_i | i \in a\}$ and c is a set.

Before describing the cumulative-cardinal hierarchy of sets we must describe **ordinals**—supports of the construction. For that purpose we need the following elementary axioms for sets:

axioms for hereditary finite sets

1. $\emptyset$ is a set.
2. If s and a are sets then $s \cup \{a\}$ is a set.

[2] Let's note that we can replace the least fixed points in the cumulative cardinal hierarchy with the greatest fixed points and get non-well-founded sets, too.

These axioms are obviously true in the cumulative-cardinal conception. Now we can define

$$s^{+} = s \cup \{s\}$$

and we can define the class Ord (see 2 Finding paradoxial classes)

$$Ord = \cap\{C \mid \emptyset \in C \ \text{ and } \ \forall\alpha(\alpha \in C \rightarrow \alpha^{+} \in C) \ \text{ and } \ \forall s \subseteq C(s = \cup s \rightarrow s \in C)\}$$

Intuitively, ordinals grow up together with the cumulative hierarchy. Formally, the next axiom, the axiom of cumulative hierarchy, gives all classical ordinals.

the **axiom of a cumulative-cardinal hierarchy**:

There is $C : Ord \longrightarrow V$ such that:

1. $C_0 = \emptyset$
2. $C_1 = \cap\{Cl \mid \emptyset \in Cl \text{ and } \forall s, a(s, a \in Cl \rightarrow s \cup \{a\} \in Cl)\}$
3. $C_{\alpha^+} = \cap\{Cl \mid \forall S \subseteq Cl(S \preceq C_\alpha \rightarrow S \in Cl)\}$, for $\alpha > 1$
4. For limit ordinal α $C_\alpha = \cap\{Cl \mid \forall S \subseteq Cl(S \preceq C_\beta \text{ for some } \beta < \alpha \rightarrow S \in Cl)\}$
5. $V = \bigcup_{\alpha \in ORD} C_\alpha = \{S \mid \exists C_\beta(S \preccurlyeq C_\beta)\}$

The meaning of the relation $\preceq$ is standard: $X \preceq Y \leftrightarrow \exists F\ F[Y] = X$

Developing the set theory from these axioms is out of the scope of this article. Therefore, we will return to the informal cumulative-cardinal conception of set, based on basic ideas of the cardinality of collecting and the possibility to iterate greater and greater collectings. The conception inherits good properties from cardinal conception and from cumulative conception of sets. It is well known that Cantor's limitation of size conception justifies easily the replacement axiom but it cannot justify the power set axiom and the union axiom. Contrary, the iterative conception justifies the power set axiom and the union axiom easily but it cannot justify the replacement axiom (Hallett 1984, p. 199; Barwise and Moss 1996, p. 208).[3] Now we can justify all ZF axioms easily (we accept the **axiom of extensionality** as a basic property of sets and the **axiom of choice** as an axiom of a different nature which is compatible with this conception):

1. the **axiom of empty set**: $\emptyset$ is a set.
   $\emptyset \subseteq C_1 \rightarrow \emptyset \in V$
2. the **axiom of infinity**: $\omega \subseteq C_1 \rightarrow \omega \in V$.
3. the **subset axiom**: $C \subseteq s \rightarrow C$ is a set,
   Because s is a set it is dominated by some C_α. But then its subclass C is dominated by C_α too, so C is also a set.
4. the **power set axiom**: s is a set $\rightarrow P(s)$ is a set.
   Because s is a set $s \subseteq C_\alpha$. But then for $a \subseteq s$ $a \subseteq C_\alpha$, so $a \in C_{\alpha^+}$. It means that $P(s) \subseteq C_{\alpha^+}$. Therefore $P(s)$ is a set.[4]

[3] For an elaborate discussion of the cumulative and cardinal conceptions of set see Boolos (1971), Boolos (1989).

[4] Because the hierarchy C_α gives a cardinal scale we can estimate where $P(C_\alpha)$ is in the scale: $C_\alpha \prec P(C_\alpha) \preccurlyeq C_{\alpha^+}$. In the cumulative cardinal conception it is natural to postulate that the hierarchy C_α provides all cardinalities. From that postulate follows the Generalized Continuum Hypothesis.

5. the **replacement axiom**: s is a set and F is a function $\rightarrow F[s]$ is a set.
   Indeed, $F[s] \preccurlyeq s \preccurlyeq C_\alpha$, so $F[s] \preccurlyeq C_\alpha$. Therefore $F[s]$ is a set.
6. the **union axiom** s is a set $\rightarrow \cup s$ is a set.
   Because $s \subseteq C_\alpha$, every x from s is in C_α. But then $x \subseteq C_\alpha$ and $\cup a = \cup\{x \mid x \in a\} \subseteq C_\alpha$. Therefore $\cup s$ is a set.
7. the **axiom of groundedness** Every set is grounded, that is to say there is no infinite descending sequence of sets $s \ni s_1 \ni s_2 \ni \cdots$.
   This is easy to prove by the $\in$-induction on sets (the principle is valid because every C_α is the least fixed point of the collecting of a certain kind. Let every element of s be grounded. If s isn't grounded then there is an infinite sequence $s \ni s_1 \ni s_2 \ni \cdots$. Then the sequence $s_1 \ni s_2 \ni \cdots$ is infinite, too, which is impossible because by the induction hypothesis s_1 is grounded.

We have got all the axioms of the Morse–Kelley set theory. Contrary, from the Morse–Kelley axioms it would be possible to prove all CC axioms (We can show that $C_{\alpha^+} = H(|C_\alpha|)$, where $|s|$ denotes the cardinality of the set s, and $H(\kappa)$ denotes the set of all sets of the hereditarily cardinality less or equal to κ). In this way the CC axioms provide a natural and plausibly consistent axiomatization for the Morse–Kelley set theory.

Acknowledgments We thank the anonymous reviewers who pointed to vague places in the initial manuscript.

5 Appendix

In the appendix the paradoxicality of some famous classes is shown in a uniform and in a direct way (see 1 The productivity principle).

The class of not-n-cyclic sets $R^n = \{x \mid \neg x \in^n x\}$ (where $x \in^n y \leftrightarrow \exists x_1, x_2, \ldots, x_{n-1}$ $x \in x_1 \in \cdots \in x_{n-1} \in y$).

Uniform way. Let $s \subseteq R^n$. If $s \in s$ then $s \in R^n$, so there is no n-cycle $s \in x_1 \in \cdots \in x_{n-1} \in s$. But such a cycle is just $s \in s \in s \in \cdots \in s$ (n times). Therefore, $s \notin s$. If s doesn't belong to the class R^n then there is n-cycle $s \in x_1 \in \cdots \in x_{n-1} \in s$. But then there is also n-cycle $x_{n-1} \in s \in x_1 \in \cdots \in x_{n-1}$, and it is impossible because x_{n-1} is an element of s, and so of R^n, therefore not-n-cyclic one. So the final conclusion is that $s \in R^n \setminus s$, that is to say identity is productive on R.

Direct way Suppose that R^n is a set. If $R^n \in R^n$ then it is on one side not-n-cyclic, because it belongs to R^n, and on the other side it is n-cyclic because it makes n-cycle $R^n \in R^n \in \cdots \in R^n$ (n times). Therefore $R^n \notin R^n$. So, it belongs to some n-cycle $R^n \in x_1 \in \cdots \in x_{n-1} \in R^n$ which gives n-cycle $x_{n-1} \in R^n \in x_1 \in \cdots \in x_{n-1}$. Therefore $x_{n-1} \in R^n$ is n-cyclic, and this is a contradiction.

Class NI of sets which are not isomorphic to its element.

We say that x is isomorphic to $y \leftrightarrow \exists F : Tr(x) \longrightarrow Tr(y)$ where F is a bijection and preserves the belonging, that is $a \in b \leftrightarrow F(a) \in F(b)$, and where $Tr(x) = \cap\{C \mid C$ is a transitive and $x \in C\} = \{y \mid \forall C(C$ is a transitive and $x \in C \rightarrow y \in C\}$.

Uniform way. Let $s \subseteq NI$. If $s \in s$ then s is isomorphic to its element (to itself). However, this is impossible because s as a subset of NI contains sets which are not isomorphic to its element. Therefore, $s \notin s$. If s is not in NI then it is isomorphic to its element x by an isomorphism F. From $x \in s$ results that $F(x) \in F(s) = x$, so x is isomorphic to its element $F(x)$. But this is impossible because x as an element of s is also an element of NI. So s is not isomorphic to its element. Therefore $s \in NI \setminus s$.

Direct way. Suppose that NI is a set. If $NI \in NI$ then it is isomorphic to its element (to itself), so $NI \notin NI$. But if it is not an element of NI then it is isomorphic to its element $x \in NI$, which is isomorphic to its element, and this is impossible because x belongs to NI. Therefore $NI \in NI$, and this is a contradiction.

Class WF of grounded sets .

We say that set x is grounded $\leftrightarrow$ there isn't a sequence of sets x_n, $n \in \omega$ such that $x = x_0 \ni x_1 \ni x_2 \ni \cdots$—the existence of natural numbers = finite ordinals is assumed.

Uniform way. Let $s \subseteq WF$. If $s \in s$ then $s \in WF$. This is impossible because there is a sequence $s \ni s \ni s \ni \cdots$. Therefore, $s \notin s$. If $s \notin WF$ then there is a sequence $s \ni x_1 \ni x_2 \ni \cdots$ and so there is also a sequence $x_1 \ni x_2 \ni \cdots$ from which there follows that x_1 is ungrounded. However, this is impossible because x_1 is an element of s, and so of WF. Therefore, $s \in WF \setminus s$.

Direct way. Let it $WF \in WF$. Then it is grounded, but we have a witness of its ungroundedness—a sequence $WF \ni WF \ni WF \ni \cdots$. So WF is ungrounded. But then there is an infinite sequence $WF \ni x_1 \ni x_2 \ni \cdots$, and so there is also an infinite sequence $x_1 \ni x_2 \cdots$ which means that x_1 is ungrounded. But this can't be because $x_1 \in WF$.

Šikić's class $S = \{x \mid x \notin F(x)\}$, for the surjection F on the universe (Šikić 1986):

Uniform way. For $s \subseteq S$ there is a d such that $s = F(d)$ (by surjectivity of F). We will show that d itself is a new element of S. If $d \in s = F(d)$ then d belongs to S from which it follows that $d \notin F(d)$. So, $d \notin F(d)$. But then $d \in S$. Therefore, $d \in S \setminus s$.

Direct way. Suppose that S is a set. Then, by surjectivity of F, there is a d such that $S = F(d)$. But then a condition for the belonging of d to set S is $d \in F(d) \leftrightarrow d \notin F(d)$, and this is a contradiction.

Examples of such operations are $x \mapsto \cup x$, $x \mapsto \cup \cup x, \ldots$ because for every set s $s = \cup\{s\} = \cup \cup \{\{s\}\}, \ldots$ (the union and the singleton axioms are assumed), and also $x \mapsto \cap x$, $x \mapsto \cap \cap x$ because for every set s $s = \cap\{s\} = \cap \cap \{\{s\}\}, \ldots$ (the intersection and the singleton axioms are assumed) (Šikić 1986).

References

Aczel, P. (1977). An introduction to inductive definitions. In J. Barwise (Ed.), *Handbook of mathematical logic*. Amsterdam: North-Holland.

Barwise, K. J., & Moss, L. (1996). *Vicious circles. On the mathematics of non-wellfounded phenomena*. CSLI Lecture Notes 60. CSLI Publications.

Boolos, G. (1971). The iterative conception of set. *Journal of Philosophy, 68*, 215–231.

Boolos, G. (1989). Iteration again. *Philosophical Topics, 17*, 5–21.

Cantor, G. (1883). Uber unendliche, lineare punktmannigfaltigkeiten, 5. *Mathematische Annalen, 21*, 545–586.

Cantor, G. (1899). Cantor's letter to Dedekind. English translation in (van Heijenoort, J. (Ed.). (1967). From Frege to Godel (pp. 113–117). Cambridge: Harvard university Press).

Fraenkel, A. (1927). *Zehn Vorlesungen uber die Grundlegung der Mengenlehre*. Leipzig: Taubner.

Fraenkel, A. (1928). *Einleitung in die Mengenlehre*. Berlin: Springer.

Fraenkel, A. A., Hillel, Y. B., & Levy, A. (1973). *Foundations of set theory*. Amsterdam: North-Holland.

Hallett, M. (1984). *Cantorian set theory and limitation of size*. Oxford logic guides 10. Oxford: Clarendon Press.

Russell, B. (1906). On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society, 4*(2), 29–53.

Šikić, Z. (1986). Cantors theorem and paradoxical classes. *Zeitschrift Fur Mathematische Logik Und Grundlagen der Mathematik, 32*(3), 221–226.