On the Proof-Theoretic Foundations of Set Theory

Lars Hallnäs

Abstract In this paper we discuss a proof-theoretic foundation of set theory that focusses on set definitions in an open type free framework. The idea to make Cantor's informal definition of the notion of a set more precise by saying that any given property defines a set seems to be in conflict with ordinary modes of reasoning. There is to some extent a confusion here between *extensional* perspectives (sets as collections of objects) and *intensional* perspectives (set theoretic definitions) that the central paradoxes build on. The solutions offered by Zermelo-Fraenkel set theories, von Neumann-Bernays set-class theories and type theories follow the strategy of retirement behind more or less safe boundaries. What if we revisit the original idea without making strong assumptions on closure properties of the theoretical notion of a set? That is, take the basic definitions for what they are without confusing the borders between intensional and extensional perspectives.

1 Introduction

Foundations of set theory relates to answers of the following two main questions:

(A) What is a set?
(B) What does it mean to reason with sets?

   With respect to (A) Cantor's informal definition of the notion of a set seems perfectly intuitive.

> By an "aggregate" (*Menge*) we understand any collection into a whole (*Zusammenfassung zu einem Ganzen*) M of definite and separate objects *m* of our intuition or our thought. [2, p. 85]

L. Hallnäs (✉)
University of Borås, Borås, Sweden
e-mail: Lars.Hallnas@hb.se

It is natural to think of *collection into a whole* as an *act of abstraction*. The question is how to understand this. In view of the paradoxes by Russell and others, the idea to make this more precise by saying that any given property defines a set seemed to be in conflict with intended *natural* modes of reasoning. What was wrong with this idea?

It might be an issue of confusing extensional and intensional perspectives. The idea of a set as a gathering of given objects into a whole paints a picture of sets as collections $(a, b, \ldots)$. We have given objects and we collect them into a whole by so to speak bracketing them. This extensional view of sets has a clear expression in the cumulative hierarchy. Abstracting with respect to a given property introduces a more intensional perspective, i.e., the way in which we actually define a set with the intention to capture a collection of objects.

> Russell's antinomy came as a veritable shock to those few thinkers who occupied themselves with foundational problems at the turn of the century. [4, p. 2]

There is something strange about this reaction. Why do we expect that such a, very general, more intensional characterisation will capture just sets as collections of objects in an intuitive extensional sense, i.e., as bracketing a given collection of objects? There is no reason to think that these two notions and perspectives should coincide, i.e., that the intensional characterisation would produce just *nice* sets, namely collections of given objects. It is in this respect of interest to note that the definition, i.e., the defining property $x \notin x$ of the Russell set R is a very elementary one. Its proof-theoretic behaviour can, for example, be observed already in intuitionistic propositional logic [3].

So if we accept the idea of abstraction with respect to any given defining property, i.e., full comprehension, as a foundation for set theory, we have an answer to question (A), that is, what a set is. But how should we then understand the paradoxes? The Russell paradox for instance seems to show that something is wrong with respect to question (B). The paradoxical argument builds on several basic assumptions, where one of the most important ones is the assumption that '*R is a set*' is a well-defined notion with respect to intended intuitive logical reasoning, which is a very strong assumption with respect to the given definition. So this is one way to view Russell's paradox; too strong assumptions on basic theoretical notions.

The solutions offered by Zermelo-Fraenkel set theories, von Neumann-Bernays set-class theories and type theories follow the strategy of retirement behind more or less safe boundaries (see [4]). There are several ideas about proof-theoretically founded restrictions on the comprehension scheme [5], [9]. Compare further the set theory of Fitch (see [4], [9]), the notion of a Frege structure [1] and notions of structural rules in relation to paradoxes [14].

Now what if we revisit the original idea without making strong assumptions on closure properties of the theoretical notion of a set? That is, take the basic definitions for what they are without confounding intensional and extensional perspectives.

2 Defining Sets

If we think of set definitions as abstractions λX, saying that a property, or functional expression, X defines a set, we may derive the following definitions of membership and equality for sets:

- $A \in \lambda X$ iff $X(A)$,
- $A = B$ iff $(x \in A \iff x \in B)$ for all sets x
  (i.e., $(A = \lambda X \ \& \ B = \lambda Y \implies \lambda X = \lambda Y) \iff (X(x) \iff Y(x))$) for all sets x).

In the same manner the axiomatic approach, ZF and other similar set theories, introduce axioms stating the existence of sets for certain specific *safe* defining properties, such as for example the subset property

$$x \in P(A) \text{ iff } x \text{ is a subset of } A$$

but also other types of axioms such as axioms introducing measurable cardinals and other large cardinals.

Although the axioms of power set and replacement, together with axioms of infinity (large cardinals starting with $\aleph_0$), provide for strong means to build sets following the cumulative hierarchy intuition of the universe of sets, they still represent a theory marked by withdrawal from foundational disasters to more favourable positions. It is not only matters of a first order formalization of *safe* axioms, but also from a more general intensional perspective a lack of elementary foundational principles. There is a very elementary and suggestive extensional picture through the cumulative hierarchy, but this is lacking with respect to definitional issues.

Why is $(x = x)$, for example, not an admissible set defining condition?

1. It contradicts the idea of sets as collections of given objects, i.e., $\lambda(x = x)$ is a member of $\lambda(x = x)$.
2. We cannot comprehend the given objects we are supposed to collect into a whole by abstraction.

In both cases we say that $(x = x)$ does not define a *set* in the sense of a *total* object that behaves nicely with respect to the intended reading of logical constants and the notion of membership. But this does not really answer the question. It just says that whatever $\lambda(x = x)$ may define it is not a set in the *extensional* sense as a collection of given objects.

The problem here is an example of what we in many cases meet as we try to define a notion where it is difficult to map out the exact borders by elementary means, the notion of a total computable function being a canonical example. From a foundational and theoretical point of view it would be nice if it were possible to make sense in some way of the initial, and very elementary, ideas of Frege and others [4].

Let us look at a very naïve and simplistic attempt to *define* sets based on the idea of sets as introduced by abstraction of defining properties. In defining sets this way it is natural to make a distinction between set expressions, i.e., sets, terms etc., and propositional expressions, i.e., propositions, formulas etc. But if we accept more open definitions this does not seem necessary, and for reasons of simplicity we will just make a distinction between sets (A) and set theoretical reasoning (B) in what follows. This would also be in line with reading *Ockham's razor* as saying that basic classifications and distinctions are matters of proofs and not foundational definitions. The definition of sets is:

- T and F are sets,
- $A \rightarrow B$, $A \in B$, $A = B$ are sets if A and B are sets,
- $S(f)$, $\exists(f)$ and $\forall(f)$ are sets if the world of sets is closed under the given function f.

This would answer question (A). To answer question (B) we add the following *derived* definition:

- T is true,
- $A \rightarrow B$ is true if (A is true $\implies B$ is true),
- $A \in S(f)$ is true if $f(A)$ is true,
- $A = B$ is true if ($x \in A$ is true $\iff x \in B$ is true) for all sets x,
- $\exists(f)$ is true if $f(x)$ is true for some set x,
- $\forall(f)$ is true if $f(x)$ is true for all sets x.

Russell's paradox tells us of course directly that there are no such definitions satisfying the intended closure properties we have written down above. But from an intensional, i.e., definitional, point of view, we actually intend to define something by writing down these clauses. The question is just what that is, in what ways we can interpret these acts of defining?

3 Functional Closure, Local Logic and the Notion of Absoluteness

There are three major issues to observe in the *definitions* given above in Sect. 2:

1. We introduce functional constructions, $S(f)$, $\exists(f)$ and $\forall(f)$, by a defining condition asking for the notion we define to be closed under a given function.
2. We introduce a conditional construction $A \rightarrow B$ by a defining condition asking B to follow from A.
3. What we actually state in the 'definitions' are closure conditions for notions we hope to be able to define in one way or another.

3.1 The Functional Closure

The idea of *function closure* (in the realm of monotone inductive definitions) is that we have some things $a, b, \ldots$ given and also some functions $f, g, \ldots$. We then define a notion X by saying that

- $a, b, \ldots$ is an X,
- if x is an X, then $f(x), g(x), \ldots$ is an X.

Implicitly this means that X is defined by these clauses and nothing else. From an intensional and foundational point of view in 'generating' X the things that $f, g, \ldots$ act on in X are not given, besides the initial things $a, b, \ldots$, they are introduced as we build X. Once defined, X is then the smallest collection of things including $a, b, \ldots$ and being closed under $f, g, \ldots$.

Similarly the idea of a functional closure is that we have some things $a, b, \ldots$ and functions $f, g, \ldots$ given and also functionals $F, G, \ldots$. Analogously, from an intensional and foundational point of view, the functions 'in' X that $F, G, \ldots$ act on, i.e., functions that X is closed under, are not given, but introduced as we build X. In both cases we take for granted certain things as primitive notions. In the first case some given objects and functions and in the second case some given objects (not necessary in all cases), some functions and functionals. In both cases what we rely on is, so to speak, inscribed in fundamental circles of reasoning. The objects we generate in building up the function closure are of course given in an abstract manner of speaking. The same thing holds for the functions we generate in building up the functional closure:

- $a, b, \ldots$ is an X,
- if x is an X, then $f(x), g(x), \ldots$ is an X,
- if X is closed under f, then $F(f), G(f), \ldots$ is an X.

In *non-foundational* and mathematically precise definitions we assume there is given a universe of objects, a function space and some functions and functionals defined on this universe/function space.

3.2 Local Logic

When defining $A \rightarrow B$ *is true* in terms of *if A is true, then B is true* it is really an issue what we mean by *if A is true, then B is true* as a defining condition. A reasonable interpretation of this is that what we mean to say is that B follows from A on the basis of information provided by the given definition, i.e., that we can prove B to follow from A in the local logic that the given definition implicitly defines. With respect to set theory this means that the sets we introduce, or to be more precise the set definitions we introduce, open up for reasoning relative to a *local set theoretic context*. What this could mean will be explained below.

3.3 Absoluteness

It is one thing to use *if A is true, then B is true* as a defining condition in a definition and quite another thing to state *if A is true, then B is true* as a closure condition for a given definition. In view of an analogy between models of set theoretical axioms and definitions of set theoretical concepts we might introduce the notion of *absoluteness* (cf. [8]) also in this definitional context. Whereas in the first case we compare how a set theoretical notion (formula) behaves in a model in relation to its behavior in another model, which intuitively means outside the model if the second model is the true cumulative hierarchy V, in the latter case we compare how a definitional notion/condition behaves inside the definition, in the local logic of the definition, with how it behaves outside the definition in the world of *intended* interpretation of defining conditions.

A set theory S is a pair of definitions $S\Phi$ and $TS\Phi$, following the ideas discussed above in Sect. 2, for a given collection of functions Φ. A defining condition A is *(left) absolute* (with respect to S), if for all defining conditions B

B follows from A in $TS\Phi$ iff (A is true in $TS\Phi \implies B$ is true in $TS\Phi$).

What this means is that deriving something from A in $TS\Phi$ is the same as implication. One closure condition that is generally self-evident is the following one

a is true by definition D iff there is a defining condition A in D of a true by D.

This is the basic axiom of definitional theory.

Take the Russell set $S(\lambda(x \in x \rightarrow F))$ (let us call it r) and let R be a set theory that includes this set. The set r is not (left) absolute in R. $r \rightarrow F$ is true in R, that is F follows from r in R. But whereas r is true in R, F is obviously not since it is not even defined in R. The argument follows from the basic definitional axiom together with an assumption that the local logic of the definition has a reasonable behavior with respect to the intended interpretation of involved logical constants. This argument demonstrates that negation is not an absolute notion, which from a proof-theoretic point of view would be a reasonable way to interpret the Russell paradox, i.e., falsity is an absolute notion, while negation is not.

This notion of absoluteness can further be specialized as follows:
A defining condition A is

1. *(right) absolute* (with respect to S) if

   A follows from B in $TS\Phi \iff$ (B is true in $TS\Phi \implies A$ is true in $TS\Phi$),

2. *upward absolute* if

   B follows from A in $TS\Phi \implies$ (A is true in $TS\Phi \implies B$ is true in $TS\Phi$),

3. *downward absolute* if

$$A \text{ is true in } TS\Phi \implies (B \text{ is true in } TS\Phi \implies B \text{ follows from } A \text{ in } TS\Phi),$$

4. etc.

To say that a defining condition, or a set, is (left/right) absolute means that the condition, or set, with respect to local reasoning has the same meaning inside the local logic as outside it.

4 A Proof-Theoretic Interpretation

Even if we note that there are no definitions having the closure properties stated in Sect. 2 above, there is still the possibility to read these *definitions* from a more strict intensional point of view. We then look at the closure conditions as clauses in two *partial inductive definitions* ([6, 7, 13]). The idea is basically to look at *if . . ., then . . .* and *is closed under* in terms of the notion of logical consequence that defines the local logic of the definitions in question, i.e., that *if A, then B* is read as *B follows from A by the given definition.*

As a mathematical object a *(partial inductive) definition D* consists of a collection of equations

$$a = A$$

for $a \in U$ for some given universe of discourse and where A is a defining condition built up from elements in U, $\top$ and $\bot$ using constructions $\bigwedge_I$ and $\Rightarrow$. Let $D(a)$ be the collection of conditions defining a in D if there are any and $\{\bot\}$ otherwise. The *local logic* of $D, \vdash_D$, is then given by the following elementary (monotone) inductive definition

$$\Gamma, a \vdash_D a$$

$$\Gamma \vdash_D \top \qquad\qquad \Gamma, \bot \vdash_D C$$

$$\frac{\Gamma \vdash_D A_i \qquad (i \in I)}{\Gamma \vdash_D \bigwedge_I A_i} \qquad\qquad \frac{\Gamma, A_i \vdash_D C}{\Gamma, \bigwedge_I A_i \vdash_D C} (i \in I)$$

$$\frac{\Gamma, A \vdash_D B}{\Gamma \vdash_D A \Rightarrow B} \qquad\qquad \frac{\Gamma \vdash_D A \qquad \Gamma, B \vdash C}{\Gamma, A \Rightarrow B \vdash_D C}$$

$$\frac{\Gamma \vdash_D A}{\Gamma \vdash_D a} (A \in D(a)) \qquad \frac{\Gamma, A \vdash_D C \qquad (A \in D(a))}{\Gamma, a \vdash_D C}$$

The function closure with respect to $X \subset U$ and functions $f_1 \ldots f_n$ with arities $k_1 \ldots k_n$ over U, is then formally defined by the following definition

$$a = \top \quad (a \in X)$$
$$f_i(x_1 \ldots x_{k_i}) = (x_1 \ldots x_{k_i}) \quad (i \leq n)$$

$Def(D(X, f_1 \ldots f_n))$ is then the smallest set containing X and being closed under the functions $f_1 \ldots f_n$.

Similarly the *functional closure* with respect to $X \subset U$, functions $f_1 \ldots f_n$ with arities $k_1 \ldots k_n$ over U, a functional $F : [U \rightarrow U] \rightarrow U$ and a set $\Phi \subset [U \rightarrow U]$, is given by a definition $D(X, f_1 \ldots f_n, F, \Phi)$:

$$a = \top \quad (a \in X)$$
$$f_i(x_1 \ldots x_{k_i}) = (x_1 \ldots x_{k_i}) \quad (i \leq n)$$
$$F(f) = \bigwedge_U (x \Rightarrow f(x)) \quad (f \in \Phi)$$

Now we might rewrite the definitions $S\Phi$ and $TS\Phi$ in the following way:

$$
S\Phi
\begin{cases}
T = \top \\[4pt]
F = \top \\[4pt]
A \rightarrow B = \bigwedge(A, B) \\[4pt]
A \in B = \bigwedge(A, B) \\[4pt]
A = B = \bigwedge(A, B) \\[4pt]
S(f) = \bigwedge_{S\Phi}(x \Rightarrow f(x)) \\[4pt]
\exists(f) = \bigwedge_{S\Phi}(x \Rightarrow f(x)) \\[4pt]
\forall(f) = \bigwedge_{S\Phi}(x \Rightarrow f(x))
\end{cases}
$$

$$
TS\Phi
\begin{cases}
T = \top \\[4pt]
A \rightarrow B = A \Rightarrow B \\[4pt]
A \in S(f) = f(A) \\[4pt]
A = B = \bigwedge_{S\Phi}((x \in A \Rightarrow x \in B), (x \in B \Rightarrow x \in A)) \\[4pt]
\exists(f) = f(x) \quad (S\Phi) \\[4pt]
\forall(f) = \bigwedge_{S\Phi}(f(x))
\end{cases}
$$

Reading them as foundational definitions we have to accept certain notions as primitive notions; the conditions T and F, the function $\rightarrow$, the functionals S, $\exists$ and $\forall$, the notion of a function and indexing families over the sets we define. In principle what amounts to understanding the functional closure as a primitive foundational notion. The resulting *formal* systems, defining the local logics of the definitions, are consequently formal systems in an informal sense. They define what a proof is as a foundational notion, providing a proof-theoretic foundation of set theory, that is, using proof-theoretical notions in an abstract and open manner (cf. the notion of a *general proof theory* in [10–12]).

5 Sets

From an extensional perspective viewing sets as collections of given sets, the notion of an elementary set connects to hierarchies of what we somehow can visualize, i.e., low levels of the cumulative hierarchy. From an intensional point of view, where the act of abstraction with respect to a given defining property/function is in focus, a natural notion of an elementary set must build on characteristics of the definition. The Levy hierarchy [8] of course shows strong connections between both perspectives for ZF, but the situation here is a bit different as we look at set definitions in much more open set theories. It is for instance clear that a set such as $S(\lambda(x = x))$ is a very elementary set with respect to its defining function.

Let us say that a set

- $S(f)$ is a Φ-*set* if $S\Phi$ is closed under f, i.e., that $f(x)$ follows from x in $S\Phi$ for all sets x in $S\Phi$,
- $S(f)$ is *elementary* if it is a Φ-set for all Φ.

Both $S(\lambda(x = x))$ and $S(\lambda(x \notin x))$ are elementary sets. A simple example of a non-elementary set is $S(\lambda(x = S(\lambda(y = a))))$.

6 Foundational Issues

It is clear that consistency is not an explicit issue in the present context. Falsity (i.e., F) is by definition something that is not defined and can thus never be proved in a set theory $S\Phi$. But consistency of course relates to issues of cut elimination for sequent calculi, which relates to upward absoluteness. So assume we have a set theory S where all basic defining conditions are absolute, or at least upward absolute. From the point of view of set theoretic reasoning the sets definable in these theories are somehow 'nice' sets.

Stating that there are functions Φ with certain properties is what here corresponds to axioms of set theory, and proving or believing that the theory $S\Phi$ is absolute in some sense corresponds to defining a model for the axioms. But while the true

cumulative hierarchy V is the universe in which these models live, a general set theory $S\Omega$ with no restrictions on functions is the context in which these definitional theories $S\Phi$ live. The big difference is that V is an extensional context, i.e., the true world of pure sets, whereas $S\Omega$ is an intensional context based on a very general notion of set definitions not presupposing a rationale of welldefinedness. What set theories $S\Phi$ reflect is not inner models, but the locality of proof logics.

The idea of reduction is somehow inherent in the notion of foundations, i.e., that we build on elementary foundations. Although we evidently just walk around in ontological circles, this idea of reduction is not meaningless. A very clear and conceptually elementary model provides a reduction in the sense that we see clearly why given axioms make sense. The argument that the idea of a reduction is an illusion since the construction of the model involves all the power of the axioms themselves does not make for a strong case. It is the suggestive simplicity and clearness of the picture the model paints that is important, i.e., that we really can *see* the construction. Simplicity with respect to definitional principles builds another type of foundations; the local logic of given definitions. The foundational construction here is the functional closure interpreted as a partial inductive definition. What is important is then that we can 'see' the proofs that build the sets and the set theoretical arguments in a very elementary sense. A typical example making the difference clear is the power set $P(A) = S(PA)$ where PAx is $\forall z(z \in x \rightarrow z \in A)$. To envision $P(A)$ as a collection of given objects involves very abstract acts of visualising for large sets A. Can we *see* the set, can we trust the axiom? It is of course clear that $S(PA)$ as a set theoretical definition opens up for logical complexity in reasoning, but in this case it is a matter of visualising proofs with respect to a given definition. Can we *see* the proofs, can we trust the definition?

The definition itself is in some sense elementary, the proofs defining reasoning in theories $S\Phi$ are also elementary in some sense. Thus there is a reduction in foundations in some sense. But in actual set theoretical practice we need to trust certain closure conditions on the definitions allowing for nice forms of reasoning for what we believe to be nice theories $S\Phi$. The major challenge here is to develop set theory within the framework of theories $S\Phi$ and explore the meaning of classical set theoretical issues in this context.

Since $S\Omega$ is closed, in the sense that each definable function f is reflected in a set $S(f)$, we have the following

Theorem *V (modulo large cardinals beyond $\aleph_0$) has a definable reflection in $S\Omega$.*

References

1. Aczel, P.: Frege structures and the notions of proposition, truth and set. In: J. Barwise, H.J. Keisler, K. Kunen (eds.) The Kleene Symposium. North-Holland (1980)
2. Cantor, G.: Contributions to the Founding of the Theory of Transfinite Numbers. Dover Publications (1955)
3. Ekman, J.: Normal proofs in set theory. Ph.D. thesis, Department of Computing Science, Chalmers University of Technology (1994)
4. Fraenkel, A.A., Bar-Hillel, Y., Levy, A.: Foundations of Set Theory. North-Holland, Amsterdam (1973)
5. Hallnäs, L.: On normalization of proofs in set theory. Ph.D. Thesis, Dissertationes Mathematicae CCLXI, Warszawa (1988)
6. Hallnäs, L.: Partial inductive definitions. Theor. Comput. Sci. **87**, 115–142 (1991)
7. Hallnäs, L., Schroeder-Heister, P.: A proof-theoretic characterization of logic programming II: programs as definitions. J. Log. Comput. **1**(5), 635–660 (1991)
8. Levy, A.: A hierarchy of formulas in set theory. In: Memoirs of the American Mathematical Society, vol. 57. American Mathematical Society, Providence (1965)
9. Prawitz, D.: Natural Deduction. A Proof-Theoretical Study. Almqvist & Wiksell, Stockholm (1965)
10. Prawitz, D.: Ideas and results in proof theory. In: Fenstad, J.E. (ed.) Proceedings of the Second Scandinavian Logic Symposium. North-Holland (1971)
11. Prawitz, D.: Towards a general proof theory. In: Suppes, P. (ed.) Logic. Methodology and the Philosophy of Science IV. North-Holland, Amsterdam (1973)
12. Prawitz, D.: On the idea of a general proof theory. Synthese **27**, 63–77 (1974)
13. Schroeder-Heister, P.: Rules of definitional reflection. In: Proceedings of the 8th Annual IEEE Symposium on Logic in Computer Science. Los Alamitos, Montreal (1993)
14. Schroeder-Heister, P.: Paradoxes and structural rules. In: Novaes, C.D., Hjortland, O.T. (eds.) Insolubles and Consequences: Essays in honour of Stephen Read. College Publications, London (2012)