Why Sets?*

Andreas Blass[1] ** and Yuri Gurevich[2]

[1] Mathematics Department, University of Michigan, Ann Arbor, MI 48109, USA
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

Dedicated to Boaz Trakhtenbrot on the occasion of his 85th birthday.

Abstract. Sets play a key role in foundations of mathematics. Why? To what extent is it an accident of history? Imagine that you have a chance to talk to mathematicians from a far-away planet. Would their mathematics be set-based? What are the alternatives to the set-theoretic foundation of mathematics? Besides, set theory seems to play a significant role in computer science; is there a good justification for that? We discuss these and some related issues.

1 Sets in Computer Science

Quisani: I wonder why sets play such a prominent role in foundations of mathematics. To what extent is it an accident of history? And I have questions about the role of sets in computer science.

Author[3]**:** Have you studied set theory?

Q: Not really but I came across set theory when I studied discrete mathematics and logic, and I looked into Enderton's book [21] a while ago. I remember that ZFC, first-order Zermelo-Fraenkel set theory with the axiom of choice, became for all practical purposes the foundation of mathematics. I can probably reconstruct the ZFC axioms.

A: Do you remember the intuitive model for ZFC.

Q: Let me see. You consider the so-called cumulative hierarchy of sets. It is a transfinite hierarchy, so that you have levels $0, 1, \ldots, \omega, \omega + 1, \ldots$ On the level zero, you have the empty set and possibly some atoms. On any other level α you have the sets of objects that occur on levels $< \alpha$. Intuitively the process never ends. To model ZFC, you just go far enough in this hierarchy so that all axioms are satisfied. Is that correct, more or less?

* This is a revised version of an article originally published in the Bulletin of the European Association for Theoretical Computer Science, Number 84, October 2004, and republished here with permission of the Association.

** Blass was partially supported by NSF grant DMS–0070723 and by a grant from Microsoft Research. This paper was written during a visit to Microsoft Research.

[3] As in our previous conversations with Quisani, we simplified the record of the conversation by blending the two authors into one who prefers "we" to "I".

A: More or less. ZFC is intended to describe the whole, never-ending universe of sets obtained in the cumulative hierarchy, but technically this universe is not a model because it's not a set. That's the reason for stopping at a stage where all the axioms are satisfied. (By Gödel's second incompleteness theorem, the existence of such a stage is an assumption that goes beyond ZFC, but it is a rather mild additional assumption.)

You should be careful about the phrase "far enough ... so that all axioms are satisfied" because "far enough" seems to suggest that any sufficiently large number of steps will do. But in fact, once you've got the axioms satisfied, you can't just go on for another step or two; you need to add many more levels to get the axioms satisfied again. You should stop at some level where your model, consisting of the sets created so far, has the closure properties required by the axioms.

Q: OK. Turning to computer science, I read at the Z users website [58] the following: "The formal specification notation Z (pronounced "zed"), useful for describing computer-based systems, is based on Zermelo-Fraenkel set theory and first order predicate logic." And I was somewhat surprised.

A: Were you surprised that they use the ZF system rather than ZFC, the Zermelo-Fraenkel system with the axiom of choice? As long as we consider only finite families of sets, the axiom of choice is unnecessary. That is, one can prove in ZF that, if X is a finite family of nonempty sets, then there is a function assigning to each set $S \in X$ one of its members. Furthermore, there is a wide class of statements, which may involve infinite sets, but for which one can prove a metatheorem saying that any sentence in this class, if provable in ZFC, is already provable in ZF; see [49, Sect. 1] for details. This class seems wide enough to cover anything likely to arise in computer science, even in its more abstract parts.

Q: That is an interesting issue in its own right but I was surprised by something else. Set theory wasn't developed to compute with. It was developed to be a foundation of mathematics.

A: There are many things that were developed for one purpose and are used for another.

Q: Sure. But, because set theory was so successful in foundations of mathematics, there may be an exaggerated expectation of the role that set theory can play in foundations of computer science. Let me try to develop my thought. What makes set theory so useful in foundations of mathematics? I see two key aspects. One aspect is that the notion of set is intuitively simple.

A: Well, it took time and effort to clarify the intuition about sets and to deal with set-theoretic paradoxes; see for example [27] and [36]. But we agree that the notion of set is intuitively simple.

Q: The other aspect is that set theory is very expressive and succinct: mathematics can be faithfully and naturally translated into set theory. This is extremely important. Imagine that somebody claims a theorem but you don't understand some notions involved. You can ask the claimer to define the notions more and more precisely. In the final account, the whole proof can be reduced to ZFC, and then the verification becomes mechanical.

Can sets play a similar role in computing? I see a big difference between the reduction to set theory in mathematics and in computing. The mathematicians do not actually translate their stuff into set theory. They just convince themselves that their subject is translatable.

A: Bourbaki [10] made a serious attempt to actually translate a nontrivial portion of mathematics into set theory, but it is an exception.

Q: Right. In computing, such translations have to be taken seriously. If you want to use a high-level language that is compiled to some set-theoretic engine, then a compiler should exist in real life, not only in principle. I guess all this boils down to the question whether the datatype of sets can be the basic datatype in computing. Can sets and set operations be implemented efficiently? Can other data be succinctly interpreted in set theory.

A: There has been an attempt made in this direction [55].

Q: Yes, and most people remained unconvinced that this was the way to go. Sequences, or lists, are appropriate as the basic datastructure.

A: We know one example where sets turned out to be more succinct than sequences as the basic datastructure.

Q: Tell me about it.

A: OK, but bear with us as we explain the background. We consider computations where inputs are finite structures, for example graphs, rather than strings.

Q: Every such structure can be presented as a string.

A: That is true. But we restrict attention to computing properties that depend only on the isomorphism type of the input structure. For example, given a bipartite graph, decide whether it has a matching. Call such properties *invariant queries.*

Q: Why the restriction?

A: Because we are interested in queries that are independent of the way the input structure is presented or implemented. Consider, for example, a database query. You want that the result depends on the database only and not on exactly how it is stored.

Q: Fine; what is the problem?

A: The original problem was this: Does there exist a query language L such that

(Restrained) every query that can be formulated in L is an invariant query computable in polynomial time, and

(Maximally expressive) every polynomial-time computable invariant query can be formulated in L.

Q: How can one ensure that all L-queries are invariant?

A: Think about first-order logic as a query language. Every first-order sentence is a query. First-order queries are *pure* in the sense that they give you no means to express a property of the input structure that is not preserved by isomorphisms. Most restrained languages in the literature are pure in that same sense.

Q: But, in principle, can a restrained language allow you to have non-invariant intermediate results? For example, can you compute a particular matching, throw away the matching and return "Yes, there is a matching"?

A: Yes, a restrained language may have non-invariant intermediate results. In fact, Ashok Chandra and David Harel, who raised the original problem in [11], considered Turing machines M that are invariant in the following sense: If M accepts one string representation of the given finite structure then it accepts them all. They asked whether there is a decidable set L of invariant polynomial time Turing machines such that, for every invariant polynomial time Turing machine T_1, there is a machine $T_2 \in L$ that computes the same query as T_1 does. In the case of a positive answer, such an L would be restrained and maximally expressive.

Q: Hmm, a decidable set of Turing machines does not look like a language.

A: One of us conjectured [31] that there is no query language, even as ugly as a decidable set of Turing machines, that is restrained and maximally expressive.

Q: But one can introduce, I guess, more and more expressive restrained languages.

A: Indeed. In particular, the necessity to deal with invariant database queries led to the introduction of a number of restrained query languages [1] including the polynomial-time version of the language $\texttt{while}_{\texttt{new}}$. In [8], Saharon Shelah and the two of us proposed a query language, let us call it BGS, that is based on set theory. BGS is pure in the sense discussed above. A polynomial time bounded version of BGS, let us call it Ptime BGS, is a restrained query language.

Q: In what sense is BGS set-theoretic?

A: It is convenient to think of BGS as a programming language. A state of a BGS program includes the input structure A, which is finite, but the state itself is an infinite structure. It contains, as elements, all hereditarily finite sets built from the elements of A. These are sets composed from the elements of A by repeated use of the pairing operation $\{x, y\}$ and the union operation $\bigcup(x) = \{y \ : \ \exists z\,(y \in z \in x)\}$. BGS uses standard set theoretic operations and employs comprehension terms $\{t(x) \;:\; x \in r \;\wedge\; \varphi(x)\}$. In any case, to make a long story short, it turned out that Ptime BGS was more expressive than the Ptime version of the language $\mathtt{while}_{\mathtt{new}}$ that works with sequences; see [9] for details. For the purpose at hand, sets happened to be more efficient than sequences.

Q: I don't understand this. A set s can be easily represented by a sequence of its elements.

A: Which sequence?

Q: Oh, I see. You may have no means to define a particular sequence of the elements of s and you cannot pick an arbitrary sequence because this would violate the purity of BGS.

A: Right. You may want to consider all $|s|!$ different sequences of the elements of s. This does not violate the purity of BGS. But, because of the polynomial time restriction, you may not have the time to deal with $|s|!$ sequences.

On the other hand, a sequence $[a_1, a_2, \ldots, a_k]$ can be succinctly represented by a set $\{[i, a_i] \;:\; 1 \leq i \leq k\}$. Ordered pairs have a simple set-theoretic representation due to Kuratowski: $[a, b] = \{\{a, b\}, \{a\}\}$.

Q: I agree that, in your context, sets are more appropriate than sequences.

A: It is also convenient to have the datatype of sets available in software specification languages.

Q: But closer to the hardware level, under the hood so to speak, we cannot deal with sets directly. They have to be represented e.g. by means of sequences.

A: You know hardware better than we do. Can one build computers that deal with sets directly?

Q: A good question. The current technology would not support a set oriented architecture.

A: What about quantum or DNA-based computing?

Q: I doubt that these new paradigms will allow us to deal with sets directly but your guess is as good as mine.

2 Sets in Mathematics

Q: Let me return to the question why sets play such a prominent role in the foundation of mathematics. But first, let me ask a more basic question: Why do we need foundations at all? Is mathematics in danger of collapsing? Most mathematicians that I know aren't concerned with foundations, and they seem to do OK.

A: Well, you already mentioned the fact that an alleged proof can be made more and more detailed until it becomes mechanically verifiable.

Q: Yes, but I'd hope that this could be done with axioms that talk about all the different sorts of objects mathematicians use – real numbers, functions, sequences, Hilbert spaces, etc. – and that directly reflect the facts that mathematicians routinely use. What's the advantage of reducing everything to sets?

A: We see three advantages. First, people have already explicitly written down adequate axiomatizations of set theory. The same could probably be done for the sort of rich theory that you described, but it would take a nontrivial effort. Besides, new sorts of objects keep entering the mathematical world.

Second, when proving that a statement is consistent with ordinary mathematics, one only has to produce a model of set theory in which the statement is true. Without the set theoretic foundation, one would have to construct a model of a much richer theory.

Third, the reduction of mathematics to set theory means that the philosopher who wants to understand the nature of mathematical concepts needs only to understand one concept, namely sets.

By the way, if someone developed mathematics on the basis of a simple concept other than sets, then these advantages would apply to that alternative foundation also.

Q: These advantages make sense but they also show why a typical mathematician never has to use the reduction to set theory. Actually, the third advantage is not entirely clear to me; it seems that by reducing mathematics to set theory the philosopher can lose some of its semantic or intuitive content. Consider a proof that complex polynomials have roots, and imagine a set-theoretic formalization of it.

A: It's not a matter of the philosopher's understanding particular mathematical results or the intuition behind them, but rather understanding the general nature of abstract, mathematical concepts.

Q: Anyway, granting the value of a reduction of mathematics to a simple foundation, why should it be set theory? For example, since sequences are so important in computing, it's natural to ask whether they could replace sets in the foundations of mathematics. Similarly, Dijkstra has suggested that multisets, also known as bags, are a natural, basic notion and should play a more prominent role in mathematics [20].

A: Both transfinite sequences and multisets have recently been proposed as foundations for mathematics in [19], where axiomatizations are given and the basic theories developed. It is too soon to say how useful or how widely accepted these foundations will be. The axiom systems proposed in [19] can be interpreted in ZFC and vice versa, so they could be regarded as just providing an alternative view of the usual universe of sets, but such alternatives may turn out to be useful aids to the intuition and they may lead to technical simplifications in some topics (and complications in others).

2.1 Adequacy of Sets

Q: This notion of interpretations seems to be crucial for foundations. The way we use set theory as a foundation is by interpreting into it the richer theories – of real numbers, functions, sequences, Hilbert spaces, etc. – that mathematicians really work with. So if another theory, say of transfinite lists, and set theory are each interpretable in the other, then they can serve as foundations for exactly the same body of mathematics, just by composing interpretations.

A: That's right, but of course composing may lead to more complicated interpretations.

Q: The crucial point, though, is that set theory can serve, via suitable, possibly complicated interpretations, as a foundation for all of ordinary mathematics. So it seems the problem of foundations for mathematics is completely solved; in other words, the study of foundations of mathematics is dead.

A: Not so fast! There are things in ordinary mathematics that "stick out" of the set-theoretic foundation.

Q: Like what?

A: If we take "stick out" in the strong sense of not even being expressible in the usual set-theoretic framework, then category theory provides an example. One wants the categories of all groups, all topological spaces, etc., and these aren't sets.

Q: So you would need proper classes, right.

A: Actually, you'd need more, since you also want things like the category of all functors from topological spaces to groups. There have been various proposals for reformulating category theory to fit into a set-like framework, ZFC with some additional axioms, but they end up talking about the category of small groups, small topological spaces, etc., where "small" amounts to considering only things below a certain stage of the cumulative hierarchy. In one of these proposals, that of Feferman [22], results proved about, say, small groups automatically imply the same results about all groups, but there is still no category of all groups.

Q: You referred to the strong sense of "stick out", so I suppose there's a weak sense.

A: That would refer to questions that can be formulated in the ZFC context but cannot be settled on the basis of the ZFC axioms. There are a great many such questions, not only in set theory itself but in topology, algebra, and analysis; see [52] for a brief description of some examples. And sometimes even the inability of ZFC to prove certain facts depends on assumptions beyond ZFC,

Q: That last statement is confusing. Give me an example.

A: Solovay [57] proved that the following theory is consistent: ZF (without the axiom of choice) plus "all sets of real numbers are Lebesgue measurable" plus the axiom of dependent choice (a weak form of the axiom of choice that is sufficient for all "nice" results in analysis, like the countable additiviity of Lebesgue measure, but not for "pathology" like non-measurable sets). For people who won't give up the axiom of choice, he also showed the consistency of ZFC plus all *definable* sets of real numbers are Lebesgue measurable (where "definable" refers to definitions by formulas of set theory in which real numbers and ordinal numbers can appear as parameters). But his proof for these results used the assumption that ZFC is consistent with the existence of an inaccessible cardinal (a certain sort of large cardinal, whose existence cannot be proved in ZFC). And Shelah [56] proved that there is no way to eliminate the assumption about an inaccessible cardinal from Solovay's proof. So the inability of ZFC to explicitly define (even with real and ordinal parameters) a specific set and prove that it isn't measurable is established subject to an assumption about cardinal numbers that themselves go beyond what ZFC can provide.

Q: I find the stronger sort of sticking out to be more interesting, because it seems to require new foundational concepts, not just new axioms.

A: New axioms are an interesting topic too. Are they really needed? And if so, then what axioms are appropriate? And why? There's a wide-ranging discussion of such issues in the collection [26].

Q: Why is it that almost all mathematical concepts can be represented set-theoretically, and even the exception you cited, category theory, seems to stick out in a way that doesn't suggest fundamental new concepts?

A: We don't know. It might be a historical accident. That is, maybe it is just the mathematics developed by human beings until now that is (almost) covered by set theory, but not necessarily the mathematics of the future or of the inhabitants of far-away planets. Or the set-theoretic interpretability of our mathematics might be due to the structure of human brains; so the human race's future mathematics would admit a set-theoretic foundation but that of alien races might not. Or set-theoretic interpretability might be a really intrinsic property of all rigorous, mathematical thought. Or there might be other explanations; feel free to dream some up.

2.2 Non-ZF Sets

Q: Returning to the intuitive idea of sets, is ZFC still the only game in town?

A: It's the biggest game, but there are others. For example, there are theories of sets and proper classes which extend ZFC. The most prominent ones are the von Neumann-Bernays-Gödel theory (NBG) and the Morse-Kelley theory (MK). In both cases the idea is to continue the cumulative hierarchy for one more step. The collections created at that last step are called proper classes.

Q: Wait a minute! You said that ZFC is intended to describe the whole cumulative hierarchy of sets. So how can there be another step? And if there is one more step, why not two or many?

A: We admit that this extra step doesn't quite make sense philosophically, in the light of the intended meaning of "set" in the ZFC axioms, but it is convenient technically. Consider some property of sets, for example the property of having exactly three members. It is convenient to refer to the multitude of the sets with this property as a single object. If this object isn't a set then it is a proper class.

There is also a less known but rather elegant extension of ZFC due to Ackermann [2]. It uses a distinction between sets and classes, but not the same distinction as in NBG or MK. For Ackermann, what makes a class a set is not that it is small but rather that it is defined without reference to the totality of all sets. It turns out [43,51] that, despite the difference in points of view, Ackermann's set theory plus an axiom of foundation is equivalent to ZF in the sense that they prove the same theorems about sets. Lévy [43] showed how to interpret Ackermann's axioms by taking an initial segment of the cumulative hierarchy as the domain of sets and a much longer initial segment as the domain of classes.

Q: Are there set theories that contradict ZFC?

A: Yes. One is Quine's "New Foundations" (NF), named after the article [50] in which it was proposed. Another is Aczel's set theory with the anti-foundation axiom [3,6].

Quine's NF is axiomatically very simple. It has the axiom of extensionality (just as in ZF) and an axiom schema of comprehension, asserting the existence of $\{x : \varphi(x)\}$ whenever $\varphi(x)$ is a stratified formula. "Stratified" means that one can attach integer "types" to all the variables so that, if $v \in w$ occurs in $\varphi(x)$, then $\mathrm{type}(v) + 1 = \mathrm{type}(w)$, and if $v = w$ occurs then $\mathrm{type}(v) = \mathrm{type}(w)$.

Q: This looks just like simple type theory.

A: Yes, but the types aren't part of the formula; stratification means only that there exist appropriate types. The point is that this restriction of comprehension seems sufficient to avoid the paradoxes.

Q: I see that it avoids Russell's paradox, since $\neg(x \in x)$ isn't stratified, but how do you know that it avoids all paradoxes?

A: We only said it seems to avoid paradoxes. Nobody has yet deduced a contradiction in NF, but nobody has a consistency proof (relative to, say, ZFC or even ZFC with large cardinals). But Jensen [34] has shown that NF becomes consistent if one weakens the extensionality axiom to allow atoms. Rosser [53] has shown how to develop many basic mathematical concepts and results in NF. For lots of information about NF and (especially) the variant NFU with atoms, see Randall Holmes's web site [32].

Q: How does NF contradict the idea of the cumulative hierarchy?

A: The formula $x = x$ is stratified, so it is an axiom of NF that there is a universal set, the set of all sets. No such thing can exist in the cumulative hierarchy, which is never completed.

Q: And what about anti-foundation?

A: This theory is similar to ZFC, but it allows sets that violate the axiom of foundation. For example, you can have a set x such that $x \in x$; you can even have $x = \{x\}$.

Q: And you could have $x \in y \in x$ and even $x = \{y\} \wedge y = \{x\}$, right?

A: Yes, but the anti-foundation axiom imposes tight controls on these things. There is only one x such that $x = \{x\}$. Using that x as the value of both x and y you get $x = \{y\} \wedge y = \{x\}$, and this pair of equations has no other solutions. The axiom says, very roughly, that if you propose some binary relation to serve (up to isomorphism) as the membership relation in a transitive set, then, as long as it's consistent with the axiom of extensionality, it will be realized exactly once. It turns out that this axiomatic system and ZFC, though they prove quite different things, are mutually interpretable. That is, one can define, within either of the two theories, strange notions of "set" and "membership" that satisfy the axioms of the other theory.

2.3 Categories

Q: What about possible replacements for sets as the fundamental concept for mathematics? You mentioned that category theory sticks out of the standard set-theoretic framework, and I've heard people say that category theory itself could replace set theory as a foundation for mathematics. But I don't understand them. A category consists of a set (or class) of objects, plus morphisms and additional structure. So category theory presupposes the notion of set. How can it serve as a foundation by itself?

A: The idea that the objects (and morphisms) of a category must be viewed as forming a set seems to be an artifact of the standard, set-theoretic way of presenting general structures, namely as sets with additional structure. One can write down the axioms of category theory as first-order sentences and then do proofs from these axioms without ever mentioning sets (or classes).

Q: Sure, but unless you're a pure formalist, you have to wonder what these first-order sentences mean. How can you explain their semantics without invoking the traditional notion of structures for first-order logic, a notion that begins with "a non-empty *set* called the universe of discourse (or base set) ..."?

A: This seems like another artifact of the set-theoretic mind-set, insisting that the semantics of first-order sentences must be expressed in terms of sets. People understood first-order sentences long before Tarski introduced the set-theoretic definition of semantics. Think of that set-theoretic definition as representing, within set theory, a pre-existing concept of meaning, just as Dedekind cuts or Cauchy sequences represent in set theory a pre-existing concept of real number.

Q: Hmmm. I'll have to think about that. It still seems hard to imagine the meaning of a first-order sentence without a set for the variables to range over. But let's suppose, at least for the sake of the discussion, that the axioms of category theory make sense without presupposing sets. Those axioms seem much too weak to serve as a foundation; after all, they have a model with one object and one morphism.

A: That's right. For foundational purposes, one needs axioms that describe not just an arbitrary category but a category with additional structure, so that its objects can represent the entities that mathematicians study.

Q: That sounds reasonable but vague. What sort of axioms are we talking about here?

A: There have been two approaches. One is to axiomatize the category of categories and the other is to axiomatize a version of the category of sets.

Q: The first of these sounds more like a genuinely category-theoretic foundation; the second mixes categories and sets.

A: Yes, but the first has had relatively little success.

Q: Why? What's its history?

A: The idea was introduced by Lawvere in [40]. He proposed axioms, in the first-order language of categories, to describe the category of categories, and to provide tools adequate for the formalization of mathematics. But three problems arose. First, as pointed out by Isbell in his review [33], the axioms didn't quite accomplish what was claimed for them. That could presumably be fixed by modifying the axioms. But there was a second problem: Although some of the axioms were quite nice and natural, others were rather unwieldy, and there were a lot of them. As a result, it looked as if the axioms had just been rigged to simulate what can be done in set theory. That's related to the third problem: The representation of some mathematical concepts in terms of categories was done by, in effect, representing them in terms of sets and then treating sets as discrete categories (categories in which the only morphisms are the identity morphisms, so the category is essentially just its set of objects). This third point should not be over-emphasized; some concepts were given very nice category-theoretic definitions. For example, the natural number system is the so-called coequalizer of a pair of morphisms between explicitly described finite categories. But the use of discrete categories for some purposes made the whole project look weak.

Q: So what about the other approach, axiomatizing the category of sets?

A: That approach, also pioneered by Lawvere [39], had considerably more success, for several reasons. First, many of the basic concepts and constructions of set theory (and even of logic, which underlies set theory) have elegant descriptions in the language of categories; specifically, they can be described as so-called adjoint functors. In the category of sets, adjoint functors provide definitions of disjoint union, cartesian product, power set, function set (i.e., the set of all functions from X to Y), and the set of natural numbers, as well as the logical connectives and quantifiers.

Q: That covers quite a lot. What other advantages does the category of sets have – or provide?

A: There is a technical advantage, namely that the axioms admit a natural weakening that describes far more categories than just the category of sets. These categories, called topoi or toposes, resemble the category of sets in many ways (including the availability of all of the constructions listed above, except that the existence of the set of natural numbers is usually not included in the definition of topos) but also differ in interesting ways (for example, the connectives and quantifiers may obey intuitionistic rather than classical logic), and there are many topoi that look quite different from the category of sets (not only non-standard models of set theory but also categories of sheaves, categories of sets with a group acting on them, and many others). As a result, set-theoretic arguments can often be applied in topoi in order to obtain results about, for example, sheaves. These ideas were introduced by Lawvere and Tierney in [42]; see [35] and [44] for further information.

Q: I don't know what sheaves are. In any case, I care mostly about foundations, so this technical advantage doesn't do much for me. What more can the category of sets do for the foundations of mathematics?

A: One can argue that the notion of abstract set described in this category-theoretic approach is closer to ordinary mathematical practice than the cumulative hierarchy described by the Zermelo-Fraenkel axioms.

Q: What is this notion of abstract set? The ZF sets look pretty abstract to me.

A: The phrase "abstract set" refers (in this context) to abstracting from any internal structure that the elements of a set may have. A typical set in the cumulative hierarchy has, as elements, other sets, and there may well be membership relations (or more complicated set-theoretic relations) between these elements. Abstract set theory gets rid of all this. As described in [41], an abstract set "is supposed to have elements, each of which has no structure, and is itself to have no internal structure, except that the elements can be distinguished as equal or unequal, and to have no external structure except for the number of elements."

Q: How is this closer to ordinary mathematical practice than the cumulative hierarchy view of sets?

A: One way to describe the difference is that the abstract view gets rid of unnecessary structure.

Q: What unnecessary structure? Give me some examples.

A: The Kuratowski representation of ordered pairs $[a, b]$ as $\{\{a\}, \{a, b\}\}$ has the side effect that a is an element of an element of the pair. This double-element relationship is an artifact of the particular coding of pairs and is not intrinsic to the notion of ordered pair.

For another example, in any of the usual set-theoretic representations of the real numbers, the basic facts about $\mathbb{R}$ depend on information about, say, members of members of real numbers – information that mathematicians would never refer to except when giving a lecture on the set-theoretic representation of the real numbers. The abstract view discards this sort of information. Of course, some structural information is needed – unlike abstract sets, the real number system has internal structure. But the relevant structure is postulated directly, say by the axioms for a complete ordered field, not obtained indirectly as a by-product of irrelevant structure.

Q: So if an abstract-set theorist wanted to talk about a set from the cumulative hierarchy, with all the structure imposed by that hierarchy, he would include that structure explicitly, rather than relying on the hierarchy to provide it.

A: Exactly. If x is a set in the cumulative hierarchy, then one can form its transitive closure t, the smallest set containing x and containing all members of its members. Then t with the membership relation $\in$ (restricted to t) is an abstract representation of t. It no longer matters what the elements of t were, because any isomorphic copy of the structure $(t, \in)$ contains the same information and lets you recover x.

Q: I have a couple of additional questions abbout the abstract approach to the real number system. First, it seems that we're getting farther from category theory and closer to set theory, especially with the completeness axiom, which talks about arbitrary subsets of $\mathbb{R}$.

A: Sets are certainly an essential ingredient of the completeness axiom for the reals, but they can still be abstract sets; we don't need the cumulative hierarchy here. As we already mentioned, it is possible to describe in category-theoretic terms the notion of power set. So category theory, in particular the notion of adjoint functor, allows one to formulate the notion of "real number system in a topos" without importing any notion of cumulative hierarchy.

Q: My second question concerns the notion of simply *postulating* the desired properties of $\mathbb{R}$, rather than *proving* them as the traditional set-theoretic approach does. The postulational approach looks like cheating.

A: It's not a matter of postulating *rather than* proving but rather postulating *separately from* proving. What is involved here is a separation of two concerns. The first concern is saying what the real number system is; here the abstract approach says $\mathbb{R}$ is a complete ordered field (not, for example, that it is the set of Dedekind cuts, or the set of equivalence classes of Cauchy sequences, or anything of that sort). The second concern is proving that such a thing exists in suitable categories. The suitable categories here are topoi, and the existence proof for $\mathbb{R}$ would use a construction like Dedekind cuts. The work of constructing $\mathbb{R}$ doesn't disappear in the category-theoretic approach, but it is separated from the definition of what $\mathbb{R}$ is.

Category-theorists often emphasize (not just for $\mathbb{R}$ but for all sorts of other things) the distinction between "what is it?" and "how do you construct it?"

Q: Well if this category-theoretic view of abstract sets is so wonderful, why isn't everybody using it?

A: There are (at least) four answers to your question. One is a matter of history. The cumulative hierarchy view of sets has been around explicitly at least since 1930 [60], and Zermelo's part of ZFC (all but the replacement and foundation axioms) goes back to 1908 [59]. ZFC has had time to demonstrate its sufficiency as a basis for ordinary mathematics. People have become accustomed to it as the foundation of mathematics, and that includes people who don't actually know what the ZFC axioms are. There is, however, a chance that the abstract view of sets will gain ground if students learn basic mathematics from books like [41].

A second reason is the simplicity of the primitive notion of set theory, the membership predicate. Perhaps, we should say "apparent simplicity," in view of the complexity of what can be coded in the cumulative hierarchy. But still, the idea of starting with just $\in$ and defining everything else is philosophically appealing. Another way to say this is that, in developing mathematics, one certainly needs the concepts of "set" and "membership"; if everything else can be developed from just an iteration of these (admittedly a transfinite iteration), why not take advantage of it?

Third, there is a technical reason. Although topos theory provides an elegant view of the set-theoretic constructions commonly used in mathematics, serious uses of the replacement axiom don't look so nice in category-theoretic terms. (By serious uses of replacement, we mean something like the proof of Borel determinacy [45], which provably [28] needs uncountably many iterations of the power set operation.) But such serious uses are still quite rare.

Q: OK, what's the fourth answer to why people aren't using the category-theoretic view of abstract sets?

A: The fourth answer is that they *are* using this point of view but just don't realize it. Mathematicians talk about ZFC as the foundation of what they do, but in fact they rarely make explicit use of the cumulative hierarchy. That hierarchy enters into their work only as an invisible support for the structures they really use – like the complete ordered field $\mathbb{R}$. When you look at what these people actually say and write, it is entirely consistent with the category-theoretic viewpoint of abstract sets equipped with just the actually needed structure.

2.4 Functions

Q: The discussion of categories, with their emphasis on morphisms alongside objects, reminds me of a way in which functions could be considered more basic than sets.

A: More basic? "As basic" seems reasonable, if one doesn't insist on representing functions set-theoretically (using ordered pairs), but in what sense do you mean "more basic"?

Q: This came up when I was a teaching assistant for a discrete mathematics class. Sets were one of the topics, and several students had trouble grasping the idea that, for example, a thing a and the set $\{a\}$ are different, or that the empty set is one thing, not nothing. They thought of a set as a physical collection, obtained by bringing the elements together, not as a separate, abstract entity.

A: Undergraduate students aren't the only people who had such difficulties; see [36] for some relevant history. But what does this have to do with functions?

Q: Well, I found that I could clarify the problem for these students by telling them to think of a set S as a black box, where you can put in any potential element x and it will tell you "yes" if $x \in S$ and "no" otherwise. So I was explaining the notion of set in terms of functions, essentially identifying a set with its characteristic function. The black-box idea, i.e., functions, seemed to be something the students could understand directly, whereas sets were best understood via functions.

A: It seems that functions are obviously abstract, so the students aren't tempted to identify them with some concrete entity, whereas they are tempted to do that with sets.

Q: That may well explain what happened with my students.

If one takes seriously the idea of functions being more basic than sets, then it seems natural to develop a theory of functions as a foundation for mathematics. Has that been tried?

A: Yes, although sometimes the distinction between using sets and using functions as the basic notion is rather blurred.

Q: Blurred how?

A: Well, the set theory now known as von Neumann-Bernays-Gödel (NBG) was first introduced by von Neumann [47,48] in terms of functions. But he minimizes the significance of using functions rather than sets. Not only do the titles of both papers say "Mengenlehre" (i.e., "set theory") with no mention of functions, but von Neumann explicitly writes that the concepts of set and function are each easily reducible to the other and that he chose functions as primitive solely for technical simplicity.[4] And when Bernays [7] recast the theory in terms of sets

[4] Wir haben statt dem Begriffe der Menge hier den Begriff der Funktion zum Grundbegriffe gemacht: die beiden Begriffe sind ja leicht aufeinander zurückzuführen. Die technische Durchführung gestaltet sich jedoch beim Zugrundelegen des Funktionsbegriffes wesentlich einfacher, allein aus diesem Grunde haben wir uns für denselben entschieden. [48, page 676]

and classes (the form in which NBG is known today), he described his work as "a modification of a system due to von Neumann," the purpose of the modification being "to remain nearer to the structure of the original Zermelo system and to utilize at the same time some of the set-theoretic concepts of the Schröder logic and of *Principia Mathematica*." Bernays doesn't mention that the primitive concept has been changed from function to set (and class). The tone of Bernays's introduction gives the impression that the change is not regarded as a significant change in content but rather as a matter of connecting with earlier work (Zermelo, Schröder, Russell, and Whitehead) and of technical convenience (Bernays mentions a "considerable simplification" vis à vis von Neumann's system).

Q: Von Neumann claimed that functions were technically simpler than sets, and Bernays claimed the opposite?

A: Yes. Of course, the set-based system that von Neumann had in mind for his comparison may have been more complex than Bernays's system. Presumably part of Bernays's work was to make the set-based approach simpler.

By the way, Gödel [30] modified Bernays's formulation slightly; in particular, he used a single membership relation, whereas Bernays had distinguished between membership in sets and membership in classes. Gödel describes his system as "essentially due to P. Bernays and ... equivalent to von Neumann's system ...." In the announcement [29], Gödel stated his consistency result in terms of von Neumann's system.

Q: So it seems we can think of von Neumann's function-based axiom system as being in some sense the same as the set-based system now known as NBG. But are there function-based foundations that aren't just variants of more familiar set-based systems?

A: The lambda calculus [4,5] and its variations fit that description. The idea here is that one works in a world of functions, with application of a function to an argument as a primitive concept. There is also the primitive notion of lambda-abstraction; given a description of a function using a free variable v, say some meaningful expression A involving v, one can produce a term $\lambda v\, A$ (which most mathematicians would write as $v \mapsto A$), denoting the function whose value at any v is given by A. In the untyped lambda calculus, one takes the functions to be defined at all arguments. That way, one doesn't need to specify sets as the domains of the functions; every function has universal domain. The typed lambda calculus is less antagonistic to sets; its functions have certain types as their domains and codomains.

Q: I've seen that the lambda calculus is used in computer science. In particular, Church's original statement [13] of his famous thesis identified the intuitive concept of computability with definability in the lambda calculus.[5] Also, lambda

[5] Church's official formulation in [13, Sect. 7] is in terms of recursiveness rather than lambda-definability, but these were proven equivalent earlier in the paper. Much earlier in the paper, Church writes in footnote 3 that the definition can be given in two ways, and he then lists lambda-definability before recursiveness. So whether Quisani

calculus plays a major role in denotational semantics. But how does it relate to foundations of mathematics?

A: Church [12] originally intended the lambda calculus as an essential part (the other part being pure logic) of a foundational system for mathematics. The other pioneers of lambda calculus, albeit in the equivalent formulation using combinators, were Schönfinkel [54] and Curry [16,17,18], and they also had foundational objectives. Unfortunately, Church's system turned out to be inconsistent [37], and the system proposed by Curry was not strong enough to serve as a general foundation for mathematics. (Schönfinkel's system was also weak, being intended just as a formulation of first-order logic.)

Q: So this approach to foundations was a dead end.

A: Not really; the task is neither dead nor ended. The original plans didn't succeed, but there has been much subsequent work, which has succeeded to a considerable extent, and which may have more successes ahead of it. Church himself developed not only the pure lambda calculus [15] (essentially the lambda part of his earlier inconsistent system, but without the logical apparatus that led to the inconsistency) but also a typed lambda calculus [14] that is essentially equivalent to the simple theory of types but expressed in terms of functions and lambda abstraction instead of sets and membership. The typed lambda calculus also provides a good way to express the internal logic of topoi (and certain other categories) [38]. It forms the underlying framework of the system developed by Martin-Löf [46] as a foundation for intuitionistic mathematics. There is also a considerable body of work by Feferman (for example [23,24,25]) on foundational systems that incorporate versions of the lambda calculus and that have both constructive and classical aspects.

Q: So if you meet mathematicians from a far-away planet, would you expect their mathematics to be set-based?

A: Not necessarily but we wouldn't be surprised if their mathematics is set-based. We would certainly expect them to have a set theory, but it might be quite different from the ones we know, and it might not be their foundation of mathematics.

Acknowledgment

We thank Akihiro Kanamori and Jan Van den Bussche for promptly reading a draft of the original version of this paper and making helpful remarks. We also thank Nachum Dershowitz for some last-minute improvements of this revised version.

---

is right here depends on whether the footnote counts as the original statement of the thesis or whether one must wait until Sect. 7.

References

1. Serge Abiteboul, Richard Hull, and Victor Vianu, *Foundations of Databases*, Addison-Wesley (1995).
2. Wilhelm Ackermann, "Zur Axiomatik der Mengenlehre," *Math. Ann.* 131 (1956) 336–345.
3. Peter Aczel, *Non-Well-Founded Sets*, CSLI Lecture Notes 14, Center for the Study of Language and Information, Stanford Univ. (1988).
4. Henk Barendregt, *The Lambda Calculus. Its Syntax and Semantics*, Studies in Logic and the Foundations of Mathematics 103, North-Holland (1984).
5. Henk Barendregt, "The impact of the lambda calculus in logic and computer science," *Bull. Symbolic Logic* 3 (1997) 181–215.
6. Jon Barwise and Lawrence Moss, *Vicious Circles. On the mathematics of non-wellfounded phenomena,* CSLI Lecture Notes 60, Center for the Study of Language and Information, Stanford Univ. (1996).
7. Paul Bernays, "A system of axiomatic set theory – Part I," *J. Symbolic Logic* 2 (1937) 65–77.
8. Andreas Blass, Yuri Gurevich, and Saharon Shelah, "Choiceless polynomial time", *Annals of Pure and Applied Logic* 100 (1999) 141–187.
9. Andreas Blass, Yuri Gurevich, and Jan Van den Bussche, "Abstract state machines and computationally complete query languages", *Information and Computation* 174 (2002) 20-36.
10. Nicolas Bourbaki, *Elements of Mathematics: Theory of Sets*, Translation from French, Addison-Wesley (1968).
11. Ashok Chandra and David Harel, "Structure and complexity of relational queries", *J. Comput. and System Sciences* 25 (1982) 99–128.
12. Alonzo Church, "A set of postulates for the foundation of logic," *Ann. Math. (2)* 33 (1932) 346–366 and 34 (1933) 839–864.
13. Alonzo Church, "An unsolvable problem of elementary number theory," *Amer. J. Math.* 58 (1936) 345–363.
14. Alonzo Church, "A formulation of the simple theory of types," *J. Symbolic Logic* 5 (1940) 56–68.
15. Alonzo Church, *The Calculi of Lambda-Conversion,* Annals of Mathematics Studies 6, Princeton Univ. Press (1941).
16. Haskell Curry, "Grundlagen der kombinatorischen Logik," *Amer. J. Math.* 52 (1930) 509–536 and 789–834.
17. Haskell Curry, "The combinatory foundations of mathematical logic," *J. Symbolic Logic* 7 (1942) 49–64.
18. Haskell Curry and Robert Feys, *Combinatory Logic. Vol. I,* North-Holland (1958).
19. Oliver Deiser, "Orte, Listen, Aggregate," Habilitationsschrift, Freie Universität Berlin (2006).
20. Edsger W. Dijkstra, "Sets are unibags," Handwritten note EWD786, April 1981, `http://www.cs.utexas.edu/users/EWD/ewd07xx/EWD786a.PDF`.
21. Herbert Enderton, *Elements of Set Theory,* Academic Press (1977).
22. Solomon Feferman, "Set-theoretical foundations of category theory," in *Reports of the Midwest Category Seminar, III*, ed. S. Mac Lane, Lecture Notes in Mathematics 106, Springer-Verlag (1969) 201–247.
23. Solomon Feferman, "A language and axioms for explicit mathematics," in *Algebra and Logic,* ed. J. Crossley, Lecture Notes in Mathematics 450, Springer-Verlag (1975) 87–139.

24. Solomon Feferman, "Constructive theories of functions and classes," in *Logic Colloquium '78,* ed. M. Boffa, D. van Dalen, and K. McAloon, Studies in Logic and the Foundations of Mathematics 97, North-Holland (1980) 159–224.
25. Solomon Feferman, "Toward useful type-free theories, I," *J. Symbolic Logic* 49 (1984) 75–111.
26. Solomon Feferman, Harvey Friedman, Penelope Maddy, and John Steel, "Does mathematics need new axioms," *Bull. Symbolic Logic* 6 (2000) 401–446.
27. Abraham Fraenkel, Yehoshua Bar-Hillel, and Azriel Lévy, *Foundations of Set Theory*, Studies in Logic and the Foundations of Mathematics 67, North-Holland (1973).
28. Harvey Friedman, "Higher set theory and mathematical practice," *Ann. Math. Logic* 2 (1970/71) 325–357.
29. Kurt Gödel, "The consistency of the axiom of choice and of the generalized continuum hypothesis," *Proc. Nat. Acad. Sci. U.S.A.* 24 (1938) 556–557.
30. Kurt Gödel, *The Consistency of the Axiom of Choice and the Generalized Continuum Hypothesis with the Axioms of Set Theory*, Annals of Mathematics Studies 3, Princeton Univ. Press (1940).
31. Yuri Gurevich, "Logic and the challenge of computer science", in *Current Trends in Theoretical Computer Science*, ed. E. Börger, Computer Science Press (1988) 1–57.
32. Randall Holmes, *New Foundations Home Page,* `http://math.boisestate.edu/~holmes/holmes/nf.html`.
33. John Isbell, Review of [40], *Mathematical Reviews* 34 (1967) #7332.
34. Ronald Jensen, "On the consistency of a slight(?) modification of Quine's NF," *Synthèse* 19 (1969) 250–263.
35. Peter T. Johnstone, *Topos Theory*, London Math. Soc. Monographs 10, Academic Press (1977).
36. Akihiro Kanamori, "The empty set, the singleton, and the ordered pair," *Bull. Symbolic Logic* 9 (2003) 273–298.
37. Stephen Kleene and J. Barkley Rosser, "The inconsistency of certain formal logics," *Ann. Math. (2)* 36 (1935) 630–636.
38. Joachim Lambek and Philip Scott, *Introduction to Higher Order Categorical Logic,* Cambridge Studies in Advanced Mathematics 7, Cambridge Univ. Press (1986).
39. F. William Lawvere, "An elementary theory of the category of sets," *Proc. Nat. Acad. Sci. U.S.A.* 52 (1964) 1506–1511.
40. F. William Lawvere, "The category of categories as a foundation for mathematics," in *Proc. Conf. Categorical Algebra (La Jolla, CA, 1965),* Springer-Verlag (1966) 1–20.
41. F. William Lawvere and Robert Rosebrugh, *Sets for Mathematicians*, Cambridge Univ. Press (2003).
42. F. William Lawvere and Myles Tierney, "Quantifiers and sheaves," in *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 1*, Gauthier-Villars (1971) 329–334.
43. Azriel Lévy, "On Ackermann's set theory," *J. Symbolic Logic* 24 (1959) 154–166.
44. Saunders Mac Lane and Ieke Moerdijk, *Sheaves in geometry and logic. A first introduction to topos theory*, Universitext, Springer-Verlag (1994).
45. Donald A. Martin, "Borel determinacy," *Ann. Math. (2)* 102 (1975) 363–371.
46. Per Martin-Löf, "An intuitionistic theory of types: predicative part," in *Proceedings of the Logic Colloquium (Bristol, July, 1973),* ed. H. E. Rose and J. C. Shepherdson, Studies in Logic and the Foundations of Mathematics 80, North-Holland (1975) 73–118.

47. Johann von Neumann, “Eine Axiomatisierung der Mengenlehre,” *J. Reine Angew. Math.* 154 (1925) 219–240. English translation in *From Frege To Gödel. A Source Book in Mathematical Logic, 1879–1931,* ed. J. van Heijenoort, Harvard University Press (1967) 393–413.
48. Johann von Neumann, “Die Axiomatisierung der Mengenlehre,” *Math. Z.* 27 (1928) 669–752
49. Richard Platek, “Eliminating the continuum hypothesis,” *J. Symbolic Logic* 34 (1969) 219–225.
50. Willard Van Orman Quine, “New foundations for mathematical logic,” *Amer. Math. Monthly,* 44 (1937) 70–80
51. William Reinhardt, “Ackermann’s set theory equals ZF,” *Ann. Math. Logic* 2 (1970) 189–249.
52. Judith Roitman, “The uses of set theory,” *Math. Intelligencer* 14 (1992) 63–69.
53. J. Barkley Rosser, *Logic for Mathematicians,* McGraw-Hill (1953).
54. Moses Schönfinkel, “Über die Bausteine der mathematischen Logik,” *Math. Ann.* 92 (1924) 305–316.
55. Jacob T. Schwartz, Robert B. K. Dewar, Ed Dubinsky, and Edmond Schonberg, *Programming with Sets: An Introduction to SETL*, Springer-Verlag (1986).
56. Saharon Shelah, “Can you take Solovay’s inaccessible away,” *Israel J. Math.* 48 (1984) 1–47.
57. Robert Solovay, “A model of set theory in which every set of reals is Lebesgue measurable,” *Ann. Math. (2)* 92 (1970) 1–56.
58. Z users website, `http://vl.zuser.org/`.
59. Ernst Zermelo, “Untersuchungen über die Grundlagen der Mengenlehre I,” *Mathematische Annalen* 65 (1908) 261–281.
60. Ernst Zermelo, “Über Grenzzahlen und Mengenbereiche, Neue Untersuchungen über die Grundlagen der Mengenlehre,” *Fundamenta Mathematicae* 16 (1930) 29–47.