The Problem of the Consistency of "New Foundations"

Randall Holmes

September 10, 2013

Abstract: I will explain the nature of the long-standing problem of the consistency of the set theory New Foundations proposed by the philosopher W. v. O. Quine in 1937 both in the prior context of the development of set theory, its usefulness in mathematics, and the problem of the "paradoxes" of set theory, and in the posterior context of partial solutions to the consistency problem and related results. I do claim to have solved this problem (this is not generally agreed yet) but I am not going to talk about that on this occasion. The talk should be accessible to a general audience of mathematicians; I hope that a graduate student or mature undergraduate would get something out of it too.

Plan of the talk

I'm going to talk about the problem of the consistency of Quine's set theory "New Foundations", which is the central issue of my tiny area of set theory.

I currently believe that I have solved this problem, but this has nothing to do with the present talk, or very little.

What I propose to do is explain what the problem is and put it in some kind of context.

Why does one need a set theory? Why is there a problem of consistency of set theories? What is New Foundations anyway and why is there a problem with it in particular? What are the relevant related results and partial solutions to the problem?

What is a set?

When I tell a layman with no maths that I work in set theory, they ask “What is a set?”. This is not a bad question. Of course, officially it has no answer, as *set* is a primitive notion in our foundations of mathematics. But thinking about the question is worthwhile. It is a good question to think about when teaching Math 187, for example.

Very likely the original appearance of sets in mathematics is in the form of geometrical figures. But the Greeks, who at least officially started the game of pure mathematics with Euclidean geometry, strongly resisted the idea that a line (for example) is a set of points, because such a set would clearly have to be infinite. For us, the difficulty is to figure out what the ancient Greeks thought a line was, if they didn’t think it was a set.

Parts and wholes

One thing that a set is pretty clearly *not* is a whole with its parts as elements.

In geometry one *could* take this view. Every geometric figure could be viewed as the set of points included in it. But then it would be the case that every point P (being itself a geometric figure) would be identical with the set $\{P\}$. Also, this picture only really works as long as only points are capable of being elements.

If a set is a whole with its elements as parts, then $\{x\}$ should be the same thing as x (as in the geometric view above). Now consider $\{\{1,2\}\}$: this set has one element, while its sole element $\{1,2\}$ has two elements, so they are distinct. I give this counterexample in Math 187: it makes a serious philosophical point about sets based just on our intuition of sets as finite unordered lists, with no issues about the infinite.

The relation of part to whole is transitive. But $1 \in \{1,2\} \in \{\{1,2\}\}$ does not imply that 1 belongs to $\{\{1,2\}\}$ (even if we had the strange view that $1 = \{1,2\}$, we would then have $2 \in \{1,2\} \in \{\{1,2\}\}$ and we would have $2 \neq \{1,2\}$). Whatever the relation of member to set it, it does not coincide with the relation of part to whole.

The parts of a set are its subsets, not its elements (this is the subject of an entire book by the philosopher David Lewis).

Sets as universals (properties)

Something a set *could* be is a universal, the reification of a property.

There are two ways we specify sets in practice, by listing for finite sets, and by giving a defining property, for sets which are either infinite or inconveniently large. $\{1, 2\}$ is a set. $\{n \mid n$ is prime and $n < 1000000\}$ is a set specification for a finite collection which might be viewed as inconveniently large. $\{n \mid n$ is prime and $n > 1000000\}$ is a specification of a set. Even the humble $\{1, 2\}$ can be expressed as $\{n \mid n = 1 \vee n = 2\}$.

If we are willing to say that properties which hold of exactly the same objects are the same (taking an extensional view of identity criteria for properties) then properties look remarkably like sets.

What are sets good for?

Sets are a very general kind of abstract structure. They make it possible to precisely specify objects which we can then conveniently identify with traditional mathematical structures such as the natural numbers or the real numbers or the points and lines of Euclidean space.

For example, we can declare that by 0 we mean $\{x \mid x \neq x\}$ (the empty set, more usually written $\emptyset$) and by $x + 1$ we mean $\{y \mid y = x\}$ (the singleton set $\{x\}$), so then we have defined 0, 1 = 0+1, 2=1+1, 3=2+1....

Oh yes. We say that a set I is *inductive* just in case $0 \in I$ and for all z, if $z \in I$ then $z + 1 \in I$, and we say that the set $\mathbb{N}$ of natural numbers is the set of all n with the property that n belongs to every inductive set. (This fills a hole in the Math 187 exposition: a definition of $\mathbb{N}$ as $\{0, 1, 2, \ldots\}$ is cheating...).

Implementation not revelation

There are various reasons to be suspicious of this procedure (Zermelo's original definition of the natural numbers). It does work in the sense that on very natural assumptions about sets one can proceed to translate the usual language of arithmetic and more generally the language of counting the elements of finite sets into set theory using Zermelo's natural numbers, and all the axioms of Peano arithmetic (and other natural assertions one would expect to be true) hold.

One suspicion is brought out by the fact that if we define the natural numbers in the same way but using $x+1 = x \cup \{x\}$ we get a different definition of the natural numbers which works just as well (the now usual definition due to von Neumann).

The point is that we are *implementing* the natural numbers, not revealing their true nature, when we "construct them in set theory".

Implementing the reals, the killer app

The implementation that really impressed in its day was the implementation of the real numbers as "Dedekind cuts". I'll give a brief account with no details.

We can implement ordered pairs (x, y) as $\{\{x\}, \{x, y\}\}$ (or in other ways, one could expand this line into a whole talk).

Implement positive rational numbers as pairs (m, n) of natural numbers (intended to represent $\frac{m}{n}$) which are relatively prime (thus finessing issues about equivalence classes). Define arithmetic and order relations on positive rational numbers as usual.

Define a positive real number as a set r of positive rational numbers with the following properties:

nontrivial: There is a positive rational which belongs to r and a positive rational which does not belong to r.

downward closed: If $x \in r$ and $y < x$ then $y \in r$.

the set has no largest element: For any $x \in r$, there is y such that $x < y \in r$.

The collection of such sets r has the correct properties to serve as an implementation of the positive reals. One can implement general reals as pairs (r, s) of positive reals, intended to represent $r - s$, with the restriction that at least one of r and s must be equal to 1 (again avoiding a resort to equivalence classes).

There is even a natural way to understand this implementation geometrically: the real number r is being represented by an open interval in the rationals, and open intervals in the rationals can be defined strictly by reference to the rationals and have all real lengths.

The scandal

This was an understanding of set theory which Bertrand Russell had around 1900. It was then rocked by a scandal, the episode of the "paradoxes of set theory". Paradoxes had been presented earlier by Burali-Forti and Cantor, but Russell's paradox was very simple and seemed to seriously endanger the foundation of the view of sets that I outlined above.

If we suppose explicitly that $x \in y$ denotes the relation "x is an element of y", and for every property P we can define the set $\{x \mid x$ has property $P\}$, we have laid the formal foundations of the view expressed above.

The property we should not think about is *non-self-membership*, defined by $x \notin x$.

Russell's Paradox

Consider the Russell class $R = \{x \mid x \notin x\}$.

For any x, $x \in R \leftrightarrow x \notin x$.

so in particular,

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

Oops.

The paradoxes in intellectual history

The paradoxes of set theory still attract attention from philosophers, and from enthusiastic amateurs who have derived the curious idea that the paradoxes of set theory are a continuing problem for the foundations of mathematics, or perhaps of reason.

I think the paradoxes (the more technical ones of Cantor and Burali-Forti as well as the pithy and so more accessible one of Russell) are simply . . . a mistake. They did not create even a hiccup in the progress of mathematics, or even the progress of set theoretical foundations of mathematics.

One solution

Zermelo in 1908 proposed one complete solution. His axiom of separation asserts that for any set A and property P there is a set $\{x \in A \mid x \text{ has property } P\}$. His other axioms provide some specific sets and allow some other constructions of sets.

This corresponds to a practical fact about mathematics (and perhaps about reason!): we do not talk about subcollections of **absolutely everything** defined by properties, but subcollections of collections of objects already given (natural kinds of objects, as a rule) carved out by properties.

With some enhancements, Zermelo's set theory became the modern set theory *ZFC* found in chapter 0 of many advanced mathematics books.

Another solution

Another solution is the simple theory of types. This very simple theory actually has a remarkably complex history, and really only took its modern form around 1930. But it has a similar motivation to the much more complicated theory of types presented by Russell and Whitehead in *Principia Mathematica* in the early years of the last century. The story of New Foundations starts with the simple theory of types.

The language of the simple theory of types includes the operations of first order logic and the primitive relations of equality and membership.

Objects in the simple theory of types are of different sorts, indexed traditionally by the natural numbers, though the use of the natural numbers as type indices does *not* imply a prior understanding of the natural numbers in the theory (a common confusion among philosophers and others philosophizing about this theory). The idea is that type 0 objects are *individuals* of an unspecified nature, type 1 objects are sets of individuals, type 2 objects are sets of type 1 objects, and so forth.

The way this is managed is through the grammar of the language used. $x = y$ is only grammatical if x and y have the same type. $x \in y$ is grammatical iff the type of y is the successor of the type of x. A common (but not essential) way to keep track of this is to use indices on variables to indicate their type: $x^i = y^i$; $u^j \in v^{j+1}$.

The axioms of simple type theory

The axioms are then very simple.

extensionality: Objects of the same positive type are equal iff they have the same elements.

comprehension: For any sentence $P(x^i)$ in the language of the theory of types in which the variable A^{i+1} does not occur,

$$(\exists A^{i+1}.(\forall x^i.x^i \in A^{i+1} \leftrightarrow P(x^i))$$

is an axiom. The object represented by A^{i+1} is quite naturally called $\{x^i \mid P(x^i)\}^{i+1}$.

A demonstration

We demonstrate some reasoning in the simple theory of types.

Define 0 as $\{x \mid (\forall y.y \notin x)\}$, the set of all sets with no elements.

Define $x + 1$ as $\{u \cup \{v\} \mid u \in x \wedge v \notin u\}$, the collection of all sets obtained by adding one more element to an element of x.

Notice that 0+1 is the set of all sets with one element, 1+1 is the set of all sets with two elements, 2+1 is the set of all sets with 3 elements, and so forth.

The set of all natural numbers can be defined as the set of all inductive sets, just as above, with the different definitions of zero and successor given here.

This is Frege's definition of the natural numbers, under which three is the set of all sets with three elements – with a qualification.

It is useful to note that the Zermelo definition of the natural numbers does not work in type theory because its successor operation takes x to $\{x\}$, which is one type higher. The situation for the von Neumann definition is even worse: the definition of $x+1$ as $x \cup \{x\}$ does not make sense in type theory at all. Once the naturals are defined, the reals can be defined just as we did above.

Where did all the type indices go?! – systematic ambiguity

Where did all the type labels go in my demonstration? What makes it valid reasoning in simple type theory is that types can be deduced for all the variables...once one assigns a type to the set to be constructed. Each of the sets definition of 0,1,2... works to give a set of type 2 or higher. The type 2 number 3 is the set of all (type 1) sets of three type 0 objects; the type 17 number 3 is the set of all (type 16) sets of three type 15 objects.

This reflects a fact about type theory: if we take any definition of a mathematical concept or proof of a theorem in simple type theory and raise all of the types of the variables appearing in it by a constant amount, we get a valid definition or proof at the higher type. In Principia, Russell called the analogous phenomenon for his much more complex type theory "systematic ambiguity".

The style of reasoning which is sensible for exposition in type theory is deliberately ambiguous about what type one is in, unless perhaps in a statement or definition so complicated that one needs to make sure that every variable can be assigned a type in accordance with our grammatical rules.

You can see an extensive development of mathematics in simple type theory in my Math 502 notes.

There is no consistency problem for simple type theory

It is easy to see that if we trust Zermelo set theory, we trust simple type theory. We show this by constructing a model of simple type theory in Zermelo set theory.

Let type 0 be represented by a set X. Then type 1 is represented by the power set of X (the set of all subsets of X), type 2 is represented by the power set of the power set of X, and so forth. The translation of the language of type theory into the language of set theory is straightforward: each variable of a given type is regarded as restricted to the appropriate iterated power set of X.

A technically more difficult but interesting exercise is to interpret a version of Zermelo set theory in the simple theory of types. It can be done, though in a limited way because Zermelo's theory is a bit stronger (and ZFC is much stronger).

New Foundations: the idea

Enter W. v. O. Quine. This eminent American philosopher and logician suggested a modest simplification of simple type theory in 1937.

He suggested that all the types are actually the same. The simple theory of types does not deny that the natural number 3 of type 15 is the same object as the natural number three of type 2; it does not allow one to either assert or deny this, as one cannot grammatically *say* it. Similarly, Quine suggested, looking at the fact that any theorem we can prove about objects of a certain type is reflected by theorems proved in the same way about each higher type, that all of these statements are in fact the same statement about the same objects. He compared the world of the simple theory of types to a hall of mirrors, and suggested that all the reflections at different types should be identified...

This idea requires a formal implementation.

New Foundations, the formal specification

New Foundations is a theory in the language of first-order logic with equality and membership. All objects are of the same sort (there are no types).

The axioms are the following:

extensionality: Objects of the same positive type are equal iff they have the same elements.

comprehension: For any sentence $P(x)$ in the language of set theory which supports an assignment of types to its variables which would make it a grammatical sentence of simple type theory, the set $\{x \mid P(x)\}$ exists.

Stratification

It is traditional to state the comprehension axiom in a way which does not depend on the language of another theory.

Definition: A formula ϕ in the language of set theory is said to be *stratified* iff there is a map σ from variables to natural numbers with the property that for any subformula $x = y$ of ϕ we have $\sigma(x) = \sigma(y)$ and for any subformula $x \in y$ of ϕ we have $\sigma(x) + 1 = \sigma(y)$.

Axiom of Stratified Comprehension: For any stratified formula $P(x)$, the set $\{x \mid P(x)\}$ exists.

Note that this is actually exactly equivalent to saying that $P(x)$ can be recast as a grammatical formula of type theory.

It is worth noting that the axiom of stratified comprehension is equivalent to a finite collection of axioms making no mention of types even indirectly as via the notion of stratification.

New Foundations is immediately and obviously as mathematically capable as the theory of types, and syntatically much more convenient.

It also has some superficially appealing features which might make a mathematician familiar with ordinary set theory and its usual motivation a bit queasy.

The universal set is a set in New Foundations. The sets make up a Boolean algebra. The Frege natural numbers afford a convenient implementation of the natural numbers: 3 can be defined as the set of all sets with three elements.

Cardinal numbers can be defined as equivalence classes of sets under the usual set theoretical relation of being the same size (sets A and B are the same size if there is a one-to-one correspondence between their elements). Ordinal numbers can be defined as equivalence classes of well-orderings under similarity. These definitions seem to court the classic paradoxes of Cantor and Burali-Forti, but these are evaded – in interesting ways which I will not discuss here.

A smaller scandal

It is usual to add axioms of Infinity and Choice to the simple theory of types. It would seem appealing to add the same axioms to New Foundations.

In 1953, E. Specker showed that the Axiom of Choice is **false** in New Foundations. His methods are beyond the scope of this talk. It follows immediately that Infinty is a theorem – if the universe were finite, then any partition of any set would be finite, and one can always make choices from a finite partition.

The problem of the consistency of New Foundations

Up until this point, all mathematical developments in New Foundations had followed developments in the theory of types. The existence of a universal set and complements raised eyebrows but it was generally thought that this was a "safe" system. With Specker's result, the system suddenly looked much less safe, and the suspicion developed that it might actually be inconsistent, like the naive set theory used before the paradoxes were discovered.

The aim of research into the problem of the consistency of New Foundations was either to construct a model of the theory (or show that it was consistent in some more indirect way) using trusted methods (those derived from the usual set theories), or to discover an actual inconsistency. No generally accepted solution to the problem along either line exists.

A nice result of Specker

In 1962, E. Specker proved a theorem which verified Quine's informal intuition for this theory.

If ϕ is a sentence in the language of our set theory, let ϕ^+ denote the formally similar sentence in which every type index is raised by 1.

Let the Ambiguity Scheme be the collection of sentences $\phi \leftrightarrow \phi^+$ in the language of type theory.

Specker showed that simple type theory plus the ambiguity scheme is consistent if and only if New Foundations is consistent. The philosophical relationship between the two theories should be evident: the theory with ambiguity embodies the intuition that statements of the same form at different types should have the same truth value – but without actually identifying the types.

The best positive result

In 1969, R. B. Jensen showed that New Foundations with extensionality weakened to allow many atoms is consistent, and in fact consistent with Infinity and Choice. His general method of proof is worth describing: we give it in a special case (the construction given here will not support the most general results possible).

Let X_0 be any infinite set. Define X_{i+1} as the power set of X_i for each natural number i.

For any formula $P(x)$ of the language of the theory of types, there is an entirely natural way to translate it into a sentence of the usual set theory in which type i is interpreted as X_i.

One modifies this interpretation to get many interpretations of "the simple theory of types with urelements", in which each type $i+1$ contains subsets of type i plus many atoms with no elements. For any strictly increasing sequence s of natural numbers, the idea is to interpret not X_i but X_{s_i} as type i, and to interpret $x^i \in y^{i+1}$ as $x \in X_{s(i)} \wedge y \in X_{s(i)+1}$: each element of $X_{s(i+1)} - X_{s(i)+1}$ is interpreted as an urelement of type $i+1$ (with no elements).

Fix a finite set Σ of sentences of the language of type theory mentioning types $0-n$. It determines a partition of the $(n+1)$-element subsets A of natural numbers determined by the truth values of the sentences in Σ in the interpretation of type theory with urelements in which the values $s(0), \ldots, s(n)$ are exactly the elements of A.

By Ramsey's theorem, this partition of the $n+1$ element sets of natural numbers into a finite number (no more than $2^{|\Sigma|}$) of compartments has an infinite homogeneous set H, which will be the range of a strictly increasing sequence of natural numbers h. In the interpretation of simple type theory with urelements based on h, any sentence $\phi \in \Sigma$ is true iff the sentence ϕ^+ obtained by raising the types in ϕ by one is true.

By compactness, the simple theory of types with urelements is consistent with the ambiguity scheme. Specker's method of proof generalizes to this theory, so we have shown that the version of New Foundations with weak extensionality, called *NFU*, is consistent. That Infinity and Choice are consistent follows if one allows X_0 to be infinite and notes that choice holds in the usual set theory, so the interpretation of choice will hold in the version of type theory interpreted here.

Consequences of the consistency of *NFU*

Whatever the badness of NF itself is, Jensen's result shows that it is not caused by the ability to construct large sets encoded in the axiom of stratified comprehension, which holds in NFU and is seen to be consistent with Infinity and Choice. The existence of the universal set, Frege naturals, and Russell-Whitehead cardinals and ordinals (for example) is seen to be possible in a consistent and mathematically capable theory, and the apparently odd ways that NF avoids the usual paradoxes can all be seen to work perfectly by looking at a model of NFU.

NFU + Infinity + Choice is a mathematically capable theory (just as handy as the simple theory of types with infinity and choice, which is adequate for all classical mathematics outside of set theory) and lacks the inconvenience of explicit types. One does need to worry about stratification or otherwise of set definitions; I have exhibited an extensive development of standard mathematical concepts from set theory in my book using NFU, and it is manageable. Stratification can sometimes be inconvenient: as Sol Feferman has also noted, the treatment of indexed families of sets and operations on indexed families of sets can be rather baroque.

Tangled systems of cardinals

In the final slides, I describe work of my own from 1995 (not in the same notation I used then).

We work in standard set theory without choice.

A *tangled system of cardinals* is a system of cardinals κ_A indexed by nonempty finite subsets A of the natural numbers with the following properties.

Define A_- for a finite set of natural numbers as $A - \{\mathtt{min}(A)\}$. We stipulate that $2^{\kappa_A} = \kappa_{A_-}$, (where A has at least two elements) that is, the size of the power set of a set of size κ_A is κ_{A_-}.

The theory of the natural model of the theory of types with a set X as type 0 is completely determined by the cardinality of X: if we have two models with type 0 of the same size, the bijection between the types 0 of the two models can be extended in a natural way to a bijection between the elements of higher types which will respect the equality and membership relations of the two models at all types.

We further stipulate that the theory of the model of the first n types of the theory of types with base type κ_A depends only on the first n elements of the set A.

The existence of a tangled system of cardinals implies the consistency of New Foundations. In fact, I engineered this concept precisely in order to allow the emulation of Jensen's proof of NFU in a version of standard set theory yielding NF rather than NFU.

The idea is to allow a set of sentences Σ mentioning types $0 - n$ to determine a partition of the finite sets A of size $n+1$ based on the truth values of the sentences in Σ in a natural model of type theory with base type of cardinality κ_A. This partition determines a homogeneous set H. A model of as many types as you like based on a large enough finite subset of H will exhibit ambiguity for sentences in Σ. Compactness allows one to conclude consistency of the simple theory of types with ambiguity for all sentences, which Specker showed is equiconsistent with New Foundations.

There are results in the reverse direction: New Foundations and extensions of New Foundations prove partial results about the existence of systems of tangled cardinals. The consistency of NF is exactly equivalent to the existence of finite fragments of all standard sizes of a system of tangled cardinals, in a sense which can be made precise.

The weirdness of NF can thus be reduced to the possibility of certain weird situations in cardinal arithmetic in ordinary set theory, without any involvement of the existence of universal sets or other large structures.

I will remark that my immediate reaction in 1995 was to thenceforth maintain that I had no opinion as to whether *NF* was consistent or not: the situation in cardinal arithmetic just described is simply strange, and there is no philosophical reason to believe that it must be possible or impossible; it is entirely a technical issue. This is the starting point of my claimed consistency proof – which I am not talking about at this time.