FUNCTIONAL SET THEORY

Michael Barr*
Department of Mathematics and Statistics
McGill University
805 Sherbrooke St., W
Montreal, Quebec
Canada H3A 2K6
Email: barr@triples.math.mcgill.ca

June 11, 1999

1 Introduction

A recent article in *The Intelligencer* described in some detail the theory of non-well-founded sets for the purpose of being able to solve freely equations of the sort $X = A \times X$ with A fixed ([Barwise & Moss, 1991]). Of course, the empty set is a solution, but what was wanted was a non-empty solution (in fact, the largest solution, if such exists). If X were a non-empty solution and $x \in X$, then we would have $x = \langle a_1, x_1 \rangle$ for some $a_1 \in A$ and $x_1 \in X$. Here I use $\langle -, - \rangle$ to denote ordered pairs. What happens next depends to some extent on the definition of ordered pair, but let us take the common definition that $\langle u, v \rangle = \{\{u\}, \{u, v\}\}$ which is a set with one or two elements, depending on whether or not $u = v$. It seems a little complicated, but it works and I believe that the same problem would occur no matter what definition of ordered pair is taken.

Let us write $u \in^2 U$ to indicate that there is a t with $u \in t$ and $t \in U$. For example, $u \in^2 \langle u, v \rangle$ and $v \in^2 \langle u, v \rangle$. Now getting back to the equation $x = \langle a_1, x_1 \rangle$ with $a_1 \in A$ and $x_1 \in X$, we have that $x_1 \in^2 x$. Since $x_1 \in X$, it can be written as $x_1 = \langle a_2, x_2 \rangle$ with $a_2 \in A$ and $x_2 \in X$. This implies that $x_2 \in^2 x_1$. Continuing in this way, we have a sequence of elements $x_n \in X$ such that

$$\cdots x_n \in^2 x_{n-1} \in^2 \cdots \in^2 x_1 \in^2 x$$

This infinitely descending $\in$-chain is banned (by a specific axiom of "well-foundedness") from standard treatments of set theory and the main point of the Barwise and Moss paper is to describe a novel kind of non-well-founded set theory in which such infinite chains are allowed. The purpose of this note is to suggest that the whole enterprise is wrong-headed.

It is clear that the "intended solution" to the equation $X = A \times X$ is that X should be the set of all sequences

$$\langle a_0, a_1, \ldots, a_n, \ldots \rangle$$

for $a_0,\ a_1,\ \ldots,\ a_n,\ \ldots \in A$. The reason that classical set theory cannot provide this solution is not that this set of sequences does not exist as that for the standard definition of product the set $\mathrm{Seq}(A)$ does not satisfy $\mathrm{Seq}(A) = A \times \mathrm{Seq}(A)$. Even if there were some alternate definition of product that allowed this definition to be satisfied, there would probably come along some similar equation for which a solution was wanted and for which the new definition was inadequate.

The theory of non-well-founded sets solves this problem and all similar problems. It is unfortunate that such solutions exist, for their main effect is to avoid giving serious consideration to the real problem: the irrelevance of actual elements in mathematics.

This is perhaps best illustrated by a two-person game I heard described by someone recently (I have unfortunately forgotten the source): Player A begins by choosing a construction of the complex numbers. Call it $\mathbf{C}$. Player B then chooses an element $z_0 \in \mathbf{C}$. Then A chooses a $z_1 \in z_0$. B responds by choosing an element $z_2 \in z_1$. They alternate this way until the

*In the preparation of this paper, I have been assisted by grants from the NSERC of Canada and the FCAR du Québec.

player unable to choose an element loses. By the well-foundedness axiom, this is a finite game. By making choices in the construction of **C** that are clearly irrelevant from a mathematical point of view, it is clearly possible to describe a myriad of different sets, each of which might function as "the" complex numbers.

There are two (at least!) rather differing views on the relation of set theory to mathematics (from which, for the purposes of this paragraph, I exclude set theory itself). The first, which I held and supposed that mathematicians held quite generally, was that we took sets seriously and imagined that all mathematical objects were sets, that the number 3 *was* the set $\{0, 1, 2\}$, that the sine function, say, was a rather complicated set and so on. A different view (which appears, on the basis of a limited sample, to be the common one), is that the main thing set theory does is to show that there are models of these things inside these axioms. In particular, if set theory is consistent, then so are the constructions that could be made using them. (However, it goes beyond consistency, since there are lots of things that are consistent with the axioms that we do not assume because there is no model of them inside the standard axioms.) One thing we agree on is that, whatever they believe, mathematicians do not act as though they took seriously the idea that the zeta function is nothing but a rather complicated set. For the purposes of this article, it doesn't matter which view you hold. The foundations described here are in any case much closer to the actual practice of mathematics.

As an example, consider the question of what is an ordered pair. One answer is that $\langle u, v\rangle = \{\{u\}, \{u, v\}\}$ as described above. Other definitions are possible, but let us ignore that. On the one view, this *is* what an ordered pair is. On the second view, that is no more than a model to show that the specification of ordered pair by $\langle u, v\rangle = \langle u', v'\rangle$ implies $u = u'$ and $v = v'$ is consistent. One thing for certain is that the specification is the only property of ordered pair that is ever used.

If one takes sets seriously, then to solve an equation $X = A \times X$ requires demonstrating the existence of a set with that property, which is impossible with well-founded sets. If one doesn't take them seriously, then the only issue is the consistency of the assumption. The equation $X =$ the set of subsets of X can easily be shown (by Cantor's diagonal argument) to be inconsistent with any reasonable idea of sets.

Until a few years ago, I knew as much about set theory as the average mathematician, which is to say virtually nothing. Had I been asked to state the replacement axiom scheme to save my life, I would have perished in ignorance. Then a colleague went on leave and I was asked to teach a course on set theory. I did teach the course and in the process a curious thing happened. I lost all respect for set theory as a foundation for mathematics. The reason is simple. In virtually all branches of mathematics we define various kinds of structures and then define some kind of admissible homomorphisms as those that preserve the structure. This is true everywhere but in set theory. The elaborate structure exists, but we pay absolutely no attention to the functions that preserve it. Every non-empty set has its elements and the elements, unless empty, have elements and they, when non-empty, have elements in turn. We get this elaborate tree and the foundation axiom requires that whenever you follow a downward path through this tree, you eventually come up empty.

What would it mean for a function to preserve this structure? For a function $f: X \longrightarrow Y$, not only should f map elements of X to elements of Y, but also if $x \in X$, then f should map the elements of x to elements of $f(x)$ and so on. For most sets X, the only function $X \longrightarrow X$ that does this is the identity! Of course, all this sounds silly because while one is interested in the elements of a set, one is virtually never interested in the elements of the elements. To take one example, which I take from a recent paper of Colin McLarty [preprint, 1991], suppose you consider two possible definitions of the natural numbers. In the first (common) definition, 0 is defined to be the empty set and we define $n = \{0, 1, \ldots, n-1\}$ so that n has n elements. In the second definition, 0 is still defined to be the empty set, but $n = \{n - 1\}$. The two definitions of natural numbers are quite different. Is the number theory that results different? Assuming it isn't, why are we concerned with the set theory at all? See also [McLarty, to appear] in which he shows how to simplify and generalize Barwise and Etchemendy's representation of the Liar paradox, precisely forgetting about the identity of elements.

The problem in all these cases is not so much that sets have elements as that the elements are sets that have elements. There is a serious reason for this. Many constructions, for example, products and quotient structures, are carried out by building sets whose elements are sets. From there it is a short distance to assuming that all elements are sets, since that certainly gives a parsimonious foundation.

The set theory of the Zermelo-Fraenkel, Gödel-Bernays and similar axioms is reminiscent of defining vectors as n-tuples of scalars in that it introduces the basis as part of the strucutre. But one does not normally suppose that linear transofrmations preserve the bases and one does not normally suppose that function between sets preserve this elementhood structure (except at the first level). The set theory I describe here is much more like the axioms for abstract vector spaces. The motivation is the same as that of any other kind of axiomatic mathematics: to abstract away from the inessential detail, in this case the membership trees.

The analogy with vector spaces can be carried one step further. Just as bases are still useful and their existence one of the most important property of vector spaces, here too it turns out to be important to be able to introduce elements and reason with them.

One point should be made. It is tempting to say that we will consider two sets indistinguishable if there is a one one correspondence (which we call an isomorphism) between them. Stated thus, this leads to some serious problems. For example, $A \times B$ certainly is in one one correspondence with $B \times A$, but if we identify them we run into trouble when $A = B$. For then this identification is a non-trivial isomorphism of $A \times A$ with itself and there is another one, namely the identity. The solution is to identify two sets with structure not merely when there is an isomorphism between them that preserves that structure, but when there is a unique isomorphism between them that does. As for sets themselves, the only time it is safe to identify them is if they each have one element or are each empty, for in those cases, and only those, there is a unique isomorphism between them.

The approach to sets that we outline below is due to F. William Lawvere and Myles Tierney, who began to study it when they were each spending two years at Dalhousie University in 1969–71. Their axioms have been refined by many people, too numerous to mention here, but they have not changed in expressive power since Lawvere and Tierney first set them out (and, by the way, showed how to prove the independence of the continuum hypothesis from them).

1.1 A categorical solution. To a category theorist, the question of solutions to $X = A \times X$ is almost a triviality. The problem is not one of finding solutions but of characterizing the correct one.

To begin with, a product of two sets is not a set, but a set plus two functions, the product projections. More precisely, a product of A with X is a set we usually denote $A \times X$ together with two functions $p\colon A \times X \longrightarrow A$ and $q\colon A \times X \longrightarrow X$ which have the property that $y \in A \times X$ implies that there are unique $a \in A$ and $x \in X$ such that $p(y) = a$ and $q(y) = x$. It should be understood that these elements y, a and x are not elements in the usual sense; for one thing they are not themselves sets and don't have elements of their own.

The equation $X = A \times X$ can be interpreted only as asking for the existence of a set X together with two functions $p\colon X \longrightarrow A$ and $q\colon X \longrightarrow X$ that satisfy the above criterion. There is no problem in producing as many solutions as wanted (unless $A = \emptyset$). Assuming, for example, that A is non-empty and at most countable, then any infinite set X can be equipped with the necessary structural functions to make it a solution. How, from this plethora of solutions, do we pick out the one that we want, the set of all sequences of elements of A?

Let us write $p\colon Seq(A) \longrightarrow A$ for the function defined by

$$p(\langle a_0, a_1, \ldots, a_n, \ldots\rangle) = a_0$$

and $q\colon \mathrm{Seq}(A) \longrightarrow \mathrm{Seq}(A)$ for the function defined by

$$q(\langle a_0, a_1, \ldots, a_n, \ldots\rangle) = \langle a_1, a_2, \ldots, a_n, \ldots\rangle$$

I emphasize that although we use elements, these are not the same as the set theoretic elements. We will explain later what these elements are. It is sufficient here to say that they are mainly used just like ordinary elements except that they themselves are without structure and a set is not in any real sense made up of its elements. But functions can be defined by giving their values on elements, as illustrated.

Now let $A \xleftarrow{\;r\;} X \xrightarrow{\;s\;} X$ be a solution, in the sense described above, to $X = A \times X$. For $x \in X$, we let $s^n(x) = s(s(\cdots(x)\cdots))$, the nth composite of s at x. Then we define $f\colon X \longrightarrow \mathrm{Seq}(A)$ by

$$f(x) = \langle r(x), r(s(x)), r(s^2(x)), \ldots, r(s^n(x)), \ldots\rangle$$

This function f is the unique function $X \longrightarrow \mathrm{Seq}(A)$ for which the diagram

$$\begin{tikzcd}
 & X \arrow[r, "s"] \arrow[dl, "p"'] \arrow[dd, "f"] & X \arrow[dd, "f"] \\
A & & \\
 & \mathrm{Seq}(A) \arrow[ul, "r"] \arrow[r, "q"'] & \mathrm{Seq}(A)
\end{tikzcd}$$

commutes. We summarize this by saying that $\mathrm{Seq}(A)$ is the *terminal* solution to the equation.

This kind of equation does not always have a solution. For example, there is no set X that satisfies the equation $X = 2^X$, the set of subsets of X. However, the essential categorical properties needed to know that a terminal solution exists to this kind of equation, when it does, has been known for at least 30 years, under the name Special Adjoint Functor Theorem.

2 Sets

In this section, we give a set of elementary axioms for the category of sets. These axioms have the same logical and philosophical status as any other axioms for set theory. They are not based on an earlier formal system since there is no earlier formal system. They are a basis for formalization, not a result of it. Thus, although a category normally begins with a set—or class—of objects and a class of arrows and domain and codomain functions and a partial operation of composition, here we begin with the undefined terms.

Although the motivation for these axioms comes directly from categorical studies, the word "category" is not used; however, words familiar from category, such as domain and codomain are used.

These axioms are written in the familiar language of set theory. The actual undefined terms of the theory are "set", "function", "domain", "codomain", "composite" and "identity". In practice, each of these will be explained in categorical terms. One more word, "element" is also used, but will be defined.

We assume the following axioms:

Set–1. All functions, and only functions, have a domain and a codomain, which are sets.

If f is a function, we denote its domain by $\mathrm{dom}\,(f)$ and its codomain by $\mathrm{cod}\,(f)$ and write $f\colon \mathrm{dom}\,(f) \longrightarrow \mathrm{cod}\,(f)$.

Set–2. If, and only if, f and g are functions for which the domain of f is the codomain of g, there is a composite function.

This composite is denoted $f \circ g$.

Set–3. If $f \circ g$ is defined, then its domain is the domain of g and its codomain is the codomain of f.

Set–4. If, and only if, A is a set, there is a function called the identity of A and denoted id_A. It has the following properties:

(a) The domain and codomain of id_A are A.

(b) For any function f with domain f, we have $f \circ \mathrm{id}_A = f$.

(c) For any function f with codomain f, we have $\mathrm{id}_A \circ f = f$.

Set–5. If f, g and h are functions for which the domain of f is the codomain of g and the domain of g is the codomain of h, whence by Axioms Set–2 and Set–3, all of $f \circ g$, $(f \circ g) \circ h$, $g \circ h$ and $f \circ (g \circ h)$ are defined, then $(f \circ g) \circ h = f \circ (g \circ h)$.

Set–6. There is a set, denoted 1, with the property that for each set A, there is exactly one function $A \longrightarrow 1$.

Any set with this property is called a singleton or one element set. If A is any set, by an element of A, we mean a function $a\colon 1 \longrightarrow A$. If $1'$ is a different singleton, we say that the element $a\colon 1 \longrightarrow A$ is *equal to* the element $a'\colon 1' \longrightarrow A$ just in case there is a function (evidently unique) $t\colon 1 \longrightarrow 1'$ such that $a' = t \circ a$. It is easy to show that each element of A is equal to one whose domain is any of the possible singletons. Thus for convenience we will assume that all elements have domain 1. We will write $a \in A$ as usual to indicate that a is an element of A. Also, if $a \in A$ and $f\colon A \longrightarrow B$, we see that $f \circ a \in B$. When it is convenient to do so, we will denote it by $f(a)$.

If x is a variable in a formula whose type is that of an element of the set A, we write $x{:}A$ to indicate that fact.

Set–7. Let A and B be sets. Then for any first order formula $\phi(x{:}A, y{:}B)$ in our language with exactly two free variables x and y, such that $\forall\, x{:}A\, \exists!\, y{:}B\, \phi(x, y)$, there is a unique function $f\colon A \longrightarrow B$ such that $f(a) = b$ if and only if $\phi(a, b)$.

Actually, we do not need an axiom scheme of this strength and, as explained below, it is sufficient to suppose only two and a half instances of it.

Set–8. For every pair of functions $f\colon A \longrightarrow C$ and $g\colon B \longrightarrow C$ with the same codomain there is a set whose elements are the pairs $\langle a, b\rangle$ for which $f(a) = g(b)$. More formally, there is a set D and functions $p\colon D \longrightarrow C$ and $q\colon D \longrightarrow C$ such that

$$\forall\, x{:}A, y{:}B\, (f(x) = g(y) \Rightarrow \exists!\, z{:}D (p(z) = x) \wedge (q(z) = y))$$

This set is called the *fibered product* or *pullback* of f and g. It is often denoted $A \times_C B$ although that notation ignores the crucial role of the functions f and g. The functions p and q are called the **projections**. For elements $a \in A$ and $b \in B$, let $\langle a, b\rangle$ denote the unique element of $A \times_C B$ for which $p(\langle a, b\rangle) = a$ and $q(\langle a, b\rangle) = b$.

One special case is particularly interesting. If $C = 1$, then since any two functions to C with the same domain are the same, the elements of the fibered product are simply pairs $\langle a, b\rangle$ with $a \in A$ and $b \in B$. In this case, we call the fibered product simply the *product* (sometimes the cartesian product) and denote it by $A \times B$.

Suppose that A and B are sets and $A \times B$ is a product with projections $p: A \times B \longrightarrow A$ and $q: A \times B \longrightarrow B$. Suppose we are given a set C and functions $r: C \longrightarrow A$ and $s: C \longrightarrow B$. Then for any $c \in C$, there is an element $\langle r(c), s(c)\rangle \in A \times B$. This formula defines, by Set–6, a function we call $\langle r, s\rangle: C \longrightarrow A \times B$. From the definition of p and q, it can be calculated that $p \circ \langle r, s\rangle = r$ and $q \circ \langle r, s\rangle = s$.

If $f: A \longrightarrow C$ and $g: B \longrightarrow D$ are functions and if $A \xleftarrow{p} A \times B \xrightarrow{q} B$ and $C \xleftarrow{r} C \times D \xrightarrow{s} D$ are products, then there is a function $\langle f \circ p, g \circ q\rangle: A \times B \longrightarrow C \times D$ that we usually denote by $f \times g$.

Set–9. For any set A, there is a set $\mathcal{P}A$ of subsets of A.

This requires a some explanation. A function $f: U \longrightarrow A$ is called an **injection** if whenever u_1 and u_2 are distinct elements of U, $f(u_1) \neq f(u_2)$. We can readily show that in any fibered product diagram,

$$\begin{tikzcd} U' \arrow[r, "f'"] \arrow[d, "h"'] & A' \arrow[d, "g"] \\ U \arrow[r, "f"'] & A \end{tikzcd}$$

if f is an injection, so is f'.

Let $A \times \mathcal{P}A$ be a product of A and $\mathcal{P}A$ and let $p: A \times \mathcal{P}A \longrightarrow A$ and $q: A \times \mathcal{P}A \longrightarrow \mathcal{P}A$ be the projections. Then we suppose that there as an object denoted $\in_A$ and an injection $w_A: \in_A \longrightarrow A \times \mathcal{P}A$. The object is intended to denote the set of pairs $\langle a, U\rangle$ such that $a \in U$. The way we say that is that for any injection $f: U \longrightarrow A$, there is a unique function $\chi(f): 1 \longrightarrow \mathcal{P}A$ such that the diagram

$$\begin{tikzcd} U \arrow[r, "f"] \arrow[d] & A \arrow[d, "{\langle \mathrm{id}, \chi(f)\rangle}"] \\ \in_A \arrow[r, "w_A"'] & A \times \mathcal{P}A \end{tikzcd}$$

is a fibered product.

Set–10. There is a set $\mathbf{N}$ and functions zero: $1 \longrightarrow \mathbf{N}$ and succ: $\mathbf{N} \longrightarrow \mathbf{N}$ that have the property that for any object T and functions $t_0: 1 \longrightarrow T$ and $t: T \longrightarrow T$ there is a unique $f: \mathbf{N} \longrightarrow T$ such that the diagram

$$\begin{tikzcd} & \mathbf{N} \arrow[d, "f"] & \mathbf{N} \arrow[l, "\mathrm{succ}"'] \arrow[d, "f"] \\ 1 \arrow[ur, "\mathrm{zero}"] \arrow[dr, "t_0"'] & & \\ & T & T \arrow[l, "t"] \end{tikzcd}$$

commutes.

Set–11. Every surjection has a right inverse.

This means that if $f: A \longrightarrow B$ has the property that for every $b \in B$ there is an $a \in A$ with $f(a) = b$, then there is a $g: B \longrightarrow A$ such that $f \circ g = \mathrm{id}$.

2.1 Variations on these axioms

These axioms can be varied in a number of ways. In fact, that is one of the most attractive features of this approach: the ease with which different kinds of set theories can be explored by varying these axioms. We will mention here a few possibilities.

The first variation doesn't change the expressive power at all. One can omit the undefined terms "set" and "identity" entirely. We can suppose that the domain and codomain of a function are also functions such that $f \circ \mathrm{dom}\,(f) = \mathrm{cod}\,(f) \circ f = f$. In other words, we are identifying sets with their identity functions. This has the advantage of reducing the number of undefined terms. The price paid is essentially psychological. We still think of sets as primary and if theory is to follow practice, then we had better keep sets, even at the price of a less elegant formulation. It is possible that a later generation would see things differently.

The second variation arises from the fact that, as mentioned above, the axiom scheme in Axiom Set–7 is stronger than needed. The following instances of it suffice to develop set theory.

Set–7a. Let $g\colon A \longrightarrow B$ be a function and let $\phi(x,y) \colon = y = g(x)$.

In this case, the function g already satisfies the condition; uniqueness is the issue here. What this axiom means is that if $f(a) = g(a)$ for all $a \in A$, then $f = g$.

Set–7b. Let $A \xrightarrow{f} C \xleftarrow{g} B$ be a pair of functions with the same codomain. Let $A \xleftarrow{p} A \times_C B \xrightarrow{q} B$ be the fibered product. Let $r\colon D \longrightarrow A$ and $s\colon D \longrightarrow B$ be functions such that $f \circ r = g \circ s$. Let $\phi(x\colon D, y\colon A \times_C B) \colon = (\langle r(x), s(x)\rangle = y)$.

This says there is a unique $h\colon D \longrightarrow A \times_C B$ such that $p \circ h = r$ and $q \circ h = s$.

Set–7c. Suppose that $g\colon U \longrightarrow A \times B$ has the property that for $u \neq u' \in U$, $g(u) \neq g(u')$. Then we define $\phi(x\colon B, y\colon \mathcal{P}A)$ by

$$\exists p, q(U \xleftarrow{p} \mathrm{dom}\ p = \mathrm{cod}\ q \xrightarrow{q} A \text{ is a fibered product of } g \text{ and } \langle \mathrm{id}_A, x\rangle) \wedge (\chi(q) = y)$$

What this axiom says is that for any injective function $g\colon U \longrightarrow A \times B$ there is a unique function we usually call $f = \chi(g)$ (it specializes to the previous use of χ when $B = 1$) with the property that for all $b \in B$, if

$$\begin{tikzcd} U_b \arrow[r, "q_b"] \arrow[d, "p_b"'] & A \arrow[d, "\langle \mathrm{id}_A, b\rangle"] \\ U \arrow[r, "f"'] & A \times B \end{tikzcd}$$

is a fibered product, then $\chi(g) \circ b = \chi(q_b)$. It can further be shown that there is a fibered product diagram

$$\begin{tikzcd} U \arrow[r, "g"] \arrow[d] & A \times B \arrow[d, "\mathrm{id}_A \times \chi(g)"] \\ \in_A \arrow[r, "w_A"'] & A \times \mathcal{P}A \end{tikzcd}$$

Other sorts of variations involve weakening Set–6. If that is done, then so-called global elements—what we here have called elements—no longer play an important role in the theory. In their stead, we use what we might call local or variable elements. A variable is determined by a parameter T, which can be any set. A T-based element of A is function $T \longrightarrow A$. With appropriate changes in the axioms to reflect this, we get variations on the axioms which looks just like intuitionistic mathematics. Other variations of the same nature lead to forcing, models in which the reals are non-standard and so on.

3 Solving the domain equation

Solving the domain equation $X = A \times X$ is not entirely trivial from these axioms. The most important thing that has to be shown is that for each pair of sets A and B there is a set B^A of functions $A \longrightarrow B$ characterized by a one one correspondence between functions $A \longrightarrow B$ and elements of B^A. One way of doing this is by defining B^A to be the subset of $\mathcal{P}(A \times B)$ consisting of all those subsets that consist of the graphs of functions. The usual construction is indirect, but probably amounts to the same thing. Details of the second approach are found in [Barr & Wells, 1985], with the precise result coming on page 183 as Theorem 1 of Section 5.4. Many treatments of these axioms use a weaker power set axiom and assume function sets explicitly. These two approaches have the same expressive power and the choice between them is purely a matter of taste.

In the case of the equation $X = A \times X$, a solution in the terms described here is to find a pair of functions $p\colon X \longrightarrow A$ and $q\colon X \longrightarrow X$ such that $A \xleftarrow{p} X \xrightarrow{q} X$ is a product diagram. This is readily possible with the set X of sequences of A, which is defined to be the function set $A^{\mathbf{N}}$ with p and q defined as evaluation at zero and succ, respectively.

3.1 Other domain equations. In the study of theoretical computer science, it is important for theoretical reasons—in order to know that computer languages that allow recursion are consistent, for example—to be able to solve equations of the sort

$$X = A_0 + A_1 \times X + A_2 \times X^2 + \cdots + A_n \times X^n + \cdots \quad (*)$$

In this equation, $A_0, A_1, \ldots$ are fixed sets, X^n stands for the product of n copies of X and the sum of sets is the so-called disjoint union. This sum is defined by saying that a function whose domain is a sum is determined by its values on each of the components. One can show that such sums exist using the axioms we have described. And the meaning of this equation is that there is a set X that can be equipped with functions that makes it the sum described above. Axiom Set–10 in fact says that the equation

$$X = 1 + X$$

has a solution and its solution is what we call the natural numbers. Using our axioms, one can show that $(*)$ always has a solution. In fact, not only some solution, but two special solutions, called the initial and terminal solutions.

To explain these notions, we write, for a set S,

$$T(S) = A_0 + A_1 \times S + A_2 \times S^2 + \cdots + A_n \times S^n + \cdots$$

If $f\colon R \longrightarrow S$ is a function, there is a function $f^2 = f \times f\colon R \times R \longrightarrow S \times S$ and similarly $f^n\colon R^n \longrightarrow S^n$ and finally, one can define a function

$$T(f)\colon T(R) \longrightarrow T(S)$$

by using the functional definition of sums. Technically, we say that T is a *functor* on the *category* of sets. This means that it takes sets to sets, functions to functions and preserves the predicates domain, codomain, composition and identity that define sets.

Now a solution $X_0 = T(X_0)$ is called initial if, for any solution $S = T(S)$ there is a unique $f\colon X_0 \longrightarrow S$ such that under the identification of X_0 with $T(X_0)$ and S with $T(S)$, we also identify f with $T(f)$. It is a little awkward to give the full definition here, since I haven't specified what the equality actually means but there is no problem being completely explicit. Analogously, $X_1 = T(X_1)$ is a terminal solution if for every $S = T(S)$, there is a unique $g\colon S \longrightarrow X_1$ such that under the identifications, $g = T(g)$. It can be shown that initial and terminal solutions of $(*)$ (and domain equations even more complicated) exist in the set theory outlined here.

For the equation we began with $X = A \times X$, the initial solution is the empty set and the terminal solution is the set of countable sequences of A. For the equation $X = 1 + X$, the integers are the initial solution and the integers with one more element is the terminal solution. although the two sets are in one one correspondence, there is no such correspondence between them that respects the structure of $X = 1 + X$. The second is usefully thought of as the ordinal $\omega + 1$.

4 Conclusions

A particular problem with standard set theory is that sets have too much structure. This structure consists of the elements, the elements of the elements and so on down the tree. For almost all mathematical purposes, this further structure is irrelevant. It is there, however, and one of its effects is to complicate the solution of domain equations.

For example, in standard set theory, for a non-empty set A the equation $X = X \times A$ has no non-empty solution. It is possible to remedy this by allowing the structure of sets to be yet more complicated. This was explored by Barwise and Moss in their Intelligencer article. A second, very different, way to solve equations of this sort is to simplify set theory so that the elements are unstructured. This requires a different approach to such things as power sets and products, but the result is a very natural set theory whose sets have just the properties that are needed and no extraneous ones. This is done by making the primitive undefined notion not set and element, but set and function with functional composition replacing membership as the fundamental binary relation.

Acknowledgements. I would like to thank Duško Pavlovìc, Robert Seely and Charles Wells who read and criticized earlier versions of this article and contributed significantly to its final form.

References

M. Barr and C. Wells, **Toposes, Triples and Theories.** Grundlehren der math. Wissenschaften 278, Springer-Verlag, Berlin, Heidelberg, New York, 1985.

J. Barwise and L. Moss, Hypersets. Mathematical Intelligencer, **13**, no. 4 (Fall, 1991), 31–41.

C. McLarty, Anti-foundation and self-reference. To appear in Journal of Philosophical Logic.

C. McLarty, Numbers can be just what they have to. Preprint, 1991.