THE NEED FOR A RIGOROUS SET THEORY

James T. Smith
San Francisco State University

Why does mathematics need a rigorous set theory? Why doesn't the informal reasoning with elementary applications of the set concept suffice? What are the goals of a rigorous theory? Have they been attained?

Through the centuries, mathematicians occasionally used the concept of a set, or collection of elements, in mathematical arguments. However, their techniques were limited to simple applications like those common now in elementary classes. For example, suppose you're studying the function $f(x) = x(x-1)(x-2)$. You inspect its roots, and determine that it's positive on the union of an open bounded interval and an open unbounded one:

$$\{ x : f(x) > 0 \} = (0,1) \cup (2,\infty).$$

You can then use the De Morgan and distributive laws to describe the complementary set, where $f(x)$ is nonnegative, as a union of intervals, without reconsidering any details of f:

$$\begin{aligned}
\{ x : f(x) \le 0 \} &= -\{ x : f(x) > 0 \} \\
&= -(\,(0,1) \cup (2,\infty)\,) \\
&= -(0,1) \cap -(2,\infty) && \text{(by De Morgan's law)} \\
&= ((-\infty,0] \cup [1,\infty)) \cap (-\infty,2] \\
&= ((-\infty,0) \cap (-\infty,2]) \cup ([1,\infty) \cap (-\infty,2]) && \text{(by distributivity)} \\
&= (-\infty,0) \cup [1,2].
\end{aligned}$$

This kind of set manipulation gives little concern, for the following reason. It's based on logical manipulation of statements like $t \in \{x : f(x) > 0\}$ and $t \in (0,1)$, and these can be rewritten without using set notation:

$$\begin{aligned}
t \in \{x : f(x) > 0\} &\Leftrightarrow f(t) > 0 \\
t \in (0,1) &\Leftrightarrow 0 < t \;\&\; t < 1.
\end{aligned}$$

Many things that you can do with the sets that occur on the left can be done without them, but perhaps less conveniently. This type of reasoning is called *virtual* set theory: you *appear* to be working with sets, but really aren't.[1]

By the late 1800s, mathematicians had begun to use sets in much more sophisticated ways. They were manipulating *sets of sets*—for example, you can regard the De Morgan/ distributivity example as verifying that the set of unions of intervals is closed under complementation. Manipulations much more complex were used to attack hard problems

[1] *Virtual* is the opposite of *transparent*: something transparent appears *not* to be there, but really *is*.

arising in advanced calculus. Some of these ultimately required reconsideration of the fundamental ideas of calculus: exactly which functions are continuous, differentiable, integrable? Those notions all involve limits, so the study turned to investigation of basic limit principles. Mathematicians discovered that the least-upper-bound principle is especially important: every bounded set of real numbers has a least upper bound. Many limit theorems discussed in calculus texts are based on that: for example, the theorem that each bounded increasing sequence has a limit. Here again, set concepts are employed in a sophisticated way.

Mathematicians went even deeper, and reconsidered the foundations of the real number system itself. Some theories were proposed that account for all the necessary properties of that system, including the least-upper-bound principle. They're all heavily dependent on set concepts. In contrast with the De Morgan/distributivity example, there's no apparent way to remove that dependence. This led to the idea that set membership is *the* fundamental mathematical concept, that *all* parts of mathematics can be reinterpreted as parts of set theory. Mathematicians set out to develop set theory as a general framework for posing and attacking all mathematical problems. They succeeded for the most part, but encountered a serious problem.

Early set theories used the *extensionality principle* uncritically. This involves the notion of *open sentence*: a string $S(x)$ of symbols including a variable x that becomes a grammatical sentence $S(t)$ if you replace x by the name t of something in the scope of the theory. It makes sense to ask, for anything t in that scope, whether $S(t)$ is true or false. According to the extensionality principle, all open sentences $S(x)$ have extensions $E = \{x : S(x)\}$ that you can manipulate as sets. That is, each such E is an object within the scope of set theory.

Some open sentences S and objects t cause problems for the extensionality principle. The simplest ones involve self-membership. Most objects t clearly *don't* belong to themselves: numbers, for example. Even most *sets* clearly don't belong to themselves: intervals don't, and even the set of *all* intervals doesn't. Is there a set of all *sets*? If so, it *must* belong to itself. Is there a set of all *finite* sets? If so, it must *not* belong to itself, because it's infinite: it contains the sets $\{0,1\}$, $\{0,1,2\}$, $\{0,1,2,3\}, \ldots,$ for example. On the other hand, if there's a set of all *infinite* sets, that *must* belong to itself.

Now consider the open sentence $S(x) = \neg(x \in x)$, known as *Russell's predicate*. According to the extensionality principle, it should have an extension E that you can manipulate as a set; the object E should lie within the scope of set theory. Thus, it must make sense to ask whether $S(E)$ should be true or false. But consider this argument:

- $S(E)$ is the sentence $\neg(E \in E)$, which is true $\Leftrightarrow$ $E \in E$ is false;
- $E = \{x : S(x)\}$, so $E \in E$ is false $\Leftrightarrow$ $S(E)$ is false.

This means that $S(E)$ is true just when it's false. It *doesn't make sense* to ask whether $S(E)$ is true or false!

This difficulty was discovered by Bertrand Russell in 1903, while reviewing a set theory book by Gottlob Frege that had been proposed as a general foundation of mathematics. Russell's argument is a version of a popular puzzle: for example, consider the barber who shaves all the village men who don't shave themselves.

There are three ways out of Russell's contradiction:

1) give up the extensionality principle in some situations; or
2) conclude that $\neg(x \in x)$ is not a proper open sentence; or
3) give up the traditional logic of two mutually exclusive truth values and devise some other system, under which Russell's argument can't be formulated or doesn't lead to contradiction.

Alternative (3) has been studied intensively by logicians, but adopted by virtually no practicing mathematicians. The problem with alternative (2) is that other open sentences besides Russell's are equally offensive, and all these are hard to distinguish from those that seem to present no problem. Alternative (2) has been studied intensively, too, but the resulting systems for identifying properly formed open sentences are so cumbersome that virtually no practicing mathematicians use them. The most common solution is alternative (1): concede that *the extensionality principle is false in general*, and rebuild set theory within the traditional logical system using only a limited form of the principle that doesn't seem to lead to contradiction.

Still wanting to develop set theory as a foundation for mathematics, scholars have followed this guideline to set up various *axiomatic* set theories. The major problem is to limit extensionality enough to avoid contradiction, but not so much as to disable the mechanisms for recreating important parts of mathematics within set theory. Several set theories seem to meet this goal. According to the 1931 results of Kurt Gödel, however, mathematicians can never be sure: any argument that a set theory should be free from contradiction must itself be suspect in the same way.

Exercise. Derive a contradiction like Russell's, using the open sentence $\exists y\,[x \in y \ \&\ y \in x]$ in place of the self-membership predicate $x \in x$.