COMMENTS ON EDWARD NELSON’S
“INTERNAL SET THEORY:
A NEW APPROACH TO NONSTANDARD ANALYSIS”

GREGORY F. LAWLER

Immediately following the commentary below, this previously published article is reprinted in its entirety: Edward Nelson, Internal set theory: a new approach to nonstandard analysis, Bull. Amer. Math. Soc. 83 (1977), no. 6, 1165–1198.

Do infinitesimals exist? This question arises naturally in calculus when one wants to write
\[ f(x + \Delta x) \simeq f(x) + f'(x) \Delta x. \]
This equation should hold for infinitesimal \( \Delta x \) where the notation \( a \simeq b \) means that \( a - b \) is infinitesimal. Much calculus can be done without bothering to define infinitesimal and using the equation above as a working definition of \( f'(x) \).

For mathematicians, of course, this is not a satisfactory starting point for a theory. We are trained that we cannot answer a question such as, “Do infinitesimals exist?” without first defining the term infinitesimal. One of the achievements of nineteenth century mathematics was to give a rigorous foundation for the differential calculus using the idea of the limit. This is what we teach in undergraduate real analysis courses usually after the students have some experience and intuition in calculus. A necessary step in this program is to describe precisely what a “real” number is. Typically, one does this by giving the properties that we believe the real numbers have, e.g., a complete ordered field, and then constructing a set of numbers with such properties using Cauchy sequences or Dedekind cuts. The vast majority of students (and probably of mathematicians) consider the construction just a formality. Clearly, the real numbers exist and have these properties. Indeed, many courses in elementary analysis choose not to construct the reals but rather to take the existence of an ordered field as given. This is reasonable: we are implicitly assuming such an object exists, otherwise, why are we studying it?

Unfortunately, the nineteenth century rigorization of calculus was not completely rigorous because it did not answer fundamental questions about set theory and logical structures. Today the most common set theory is Zermelo–Fraenkl (ZF) usually with the axiom of choice added (ZFC). While technically this is the starting point for proofs of almost all of mathematics, few could actually give the axioms of ZF. It is also not known whether or not the axiom system ZFC is consistent! Of course, this does not really disturb most mathematicians since they “know” that...
what they do is fine. However, I think it is useful for mathematicians to realize
that existence of infinite sets and power sets are axioms. There is no proof of the
existence of an infinite set.

Given the real numbers, how do we define infinitesimal? One possibility is to say
that a number $\Delta x$ is infinitesimal if $|\Delta x|$ is less than every positive real number.
In this case, it is easy to see that 0 is the only infinitesimal and therefore most
mathematicians would say there are no nonzero infinitesimals.

Nonstandard analysis was invented by Abraham Robinson and developed in his
monograph [2]. Robinson demonstrated that one could have a consistent theory
of calculus that included infinitesimals. His construction uses the definition of
infinitesimal as in the previous paragraph. Therefore, in order to have nonzero
infinitesimals, numbers must be added to the real numbers creating an ordered field
often called the hyperreals. This construction is one case of nonstandard models.
In this case, the binary operations $+$ and $\times$ are extended to the larger set in a
way that preserves all the theorems of calculus. While this theory is beautiful, the
approach has some drawbacks: first, to even understand the construction one has
to learn a good amount of logic, and second, since the theory is based on adding
numbers to the real numbers, there is a feeling of artificiality in the construction.

We now come to this paper where Edward Nelson offers a different approach to
infinitesimals. The starting point of the theory is the assumption that infinitesimals
already exist in the real line! To understand the axioms, let us first consider the set
of positive integers. Suppose that we accept the fact that we are finite individuals
with finite capacities. Then, there should be a limit to the size of numbers we
can comprehend. Let us call a number conceivable if we can imagine a quantity
of that size. Then we would assume that 1, 2, 3, . . . are conceivable numbers. We
would also assume that there is a number $N$ such that no number greater than
$N$ is conceivable. A little more thought would lead to the assumption that if $n$
is conceivable, then so is $n + 1$. Suppose also that we would like the induction
hypothesis: if $S$ is a subset of the natural numbers containing 1 and satisfying
$n \in S$ implies $n + 1 \in S$, then $S$ is the entire set of natural numbers. We have
reached a contradiction. However, in Nelson’s theory, this contradiction is removed
by saying that the collection of conceivable positive integers does not form a set
and hence the induction property does not apply to it.

This paper introduces the precise theory which he calls Internal Set Theory (IST)
starting with the addition of a predicate standard to the usual axioms of ZFC. The
conceivable integers in the previous paragraph can be viewed as the standard in-
tegers. More generally, any real number that can be described in a conceivable
amount of time (without using the word “standard”) is standard. For example
$\pi, e, \ldots$ are standard reals. Although, we think of standard this way, in the formal
system it is an undefined predicate. The axioms of the system are labeled by the
same TLA as Internal Set Theory, IST (idealization, standardization, transfer).
These are made precise in the paper, so I will not describe them here. The critical
fact is that any theorem about the reals that we had before adding the word “stan-
dard” is still true in the new theory. Also, one of the consequences of idealization
is Theorem 1.2: there exists a finite set that contains all standard real numbers!
However, there does not exist a set that contains only the standard real numbers.

\footnote{Three-letter acronym}
One of the beauties of IST is that one does not need to learn logic in order to use it, provided one is willing to accept that ZFC with the new predicate and axioms is a conservative extension of ZFC. The proof of this theorem, which Nelson credits to William C. Powell and which uses an ultrafilter construction, is appropriately put in an appendix. I emphasize that this result does not prove that IST is consistent, but rather only that if ZFC is consistent, then IST is consistent. The construction in this proof is similar to that used by Robinson. There is nothing in the construction that a user needs to do IST. This is somewhat analogous to the fact that one does not need the Dedekind cut or Cauchy sequence construction of the real numbers to do real analysis.

The bulk of this paper is accessible to any graduate student in mathematics and could also be used in a course for advanced undergraduates. Almost all readers at some point will think they have found a contradiction to the theory—how could all of this be true? Yet, by understanding that this is all consistent, one gets a greater understanding of the real numbers. Those who enjoy this paper and want more should continue with Nelson’s monograph Radically Elementary Probability Theory, which does sophisticated probability theory using (nonstandard) finite probability spaces.

This leaves the metamathematical, or maybe philosophical, question, Do infinitesimals actually exist? The point in this paper, as well as in the work of Robinson, is that the basic theorems of analysis do not change if one assumes that they exist. The question of existence cannot be determined mathematically, and, by the nature of their “definition”, it cannot be determined empirically either. Once we start questioning, we can ask about the axioms of set theory. Can we determine empirically whether or not an infinite set exists? At least to me, the most that can be deduced from “real world observation” is that there is no set of conceivable cardinality that contains everything (even this assumption could be debated). If we look at the real line, how can we tell that it does not consist only of the numbers

$$\pm \frac{j}{N}, \quad j = 0, 1, 2, \ldots, MN,$$

where $M, N$ are unlimited integers?

Indeed, we can do all of calculus using this set of numbers provided that we are happy to use $\simeq$ as our notion of equality. However, I would not particularly recommend it. For example, we would have the numbers $\sqrt{2}, \pi$ and $e$ (There would, of course, be numbers infinitely close to them; e.g., there would be a number $x$ with $x^2 \simeq 2$.) Although taste is a personal thing, I believe that most of us would find that the traditional real number system is aesthetically nicer and easier to do calculus with. And we can be happy to know that we do not need to answer whether or not lines really have a “quantum” of $1/N$—our theorems about our fictional uncountable set $\mathbb{R}$ will still apply to the “real world”.

I think it is not so clear whether nonstandard, in particular IST, or standard methods in calculus are more aesthetic. I believe that some things are more easily stated using infinitesimals while other things are more neatly phrased in terms of the traditional reals. I have used a little nonstandard analysis in my work, but I rarely use it now at least in my written work. The main reason is that most people do not have the background. One motive in my choosing this paper is the hope that more mathematicians will learn a little about nonstandard analysis so they are not afraid when someone like me uses the word infinitesimal!
References


Department of Mathematics, University of Chicago, Chicago, Illinois 60637
E-mail address: lawler@math.uchicago.edu