INTRODUCTION

The mathematicians of the seventeenth and eighteenth centuries used a method based on the notion of an infinitesimal (an infinitely small number) to create and develop calculus and mathematical analysis. Although their method was intuitively appealing and enabled simple arguments and calculations, by the end of the nineteenth century it had to be abandoned for lack of rigor.

A rigorous development of classical analysis requires a precise definition of the real numbers, and for this the notion of an infinite set is essential. In this introduction, we provide some background for these foundational matters of mathematical analysis.

0.1 Infinite sets and the continuum

Among the most rudimentary operations of the human mind are the acts of considering a number of entities as a unit and of regarding a single object as composed of a number of constituents.[1] For example, when we use words such as *population* or *nation* we are regarding a number of entities as a unit. Similarly, when we regard a drop of water as made up of a number of molecules of water or a line segment as an assemblage of infinitely many points, we are thinking of a single object as composed of a number of constituents. The intuitive notion of a collection of elements occurs to us in conjunction with these basic mental activities.

An intellectual feat of the late nineteenth and early twentieth centuries was the discovery that virtually all known mathematics could be described in terms of ideas rooted in the intuitive notion of a "collection." This can be done once our intuitive understanding of collections and their properties acquires refinement, precision,

[1] These are essentially the same mental operations that in other contexts are referred to as *synthesizing* and *analyzing*, respectively.

and suitable expansion through a linguistic model. The word *set* is normally used in mathematics to refer to any such modified version of the intuitive notion of a collection. Indeed, such modifications have yielded sophisticated linguistic frameworks that serve as the foundations of mathematical and other scientific theories.

Each such foundational framework is called a *set theory*. There is a core common to most set theories, known as *elementary set theory*. The essential part of elementary set theory is concerned with the development of a language which, by being cast in symbols, endows with precision our intuitive understanding of the properties of collections. Most educated people today are expected to have some familiarity with elementary set theory.

More sophisticated theories of sets are obtained as we begin to expand our raw notion of a collection by assigning to it properties that may go beyond what is clearly possessed by physical or concrete examples of collections. But this transcendence entails controversies. One important controversy in the history of mathematics concerns the assumption that collections with infinitely many elements have actual (as opposed to potential) existence. One may argue, for example, that our ordinary experience of the physical world hardly informs us of the existence of such collections. Physicists tell us that even the whole universe consists of only finitely many particles.

So what makes us *think* of infinite sets? We may answer this question in one word – a continuum. We perceive the physical world in terms of time and space, which the mind grasps as continua. Herein lies the origin of geometric concepts such as straight lines, curves, planes, and surfaces. Such geometric concepts have historically been called continuous quantities. One fruitful method of studying continuous quantities is through the analytic models that we construct for them. The notion of an infinite set in its naïve form naturally occurs to us as we attempt to construct such models.

But it is one thing to be in possession of a raw idea, and quite another to define an impeccable mathematical concept that can blossom into a sophisticated theory. Indeed, in its naïve form the notion of an infinite set has been around since ancient times. But we were not in possession of a rigorous mathematical theory of infinite sets until the turn of the twentieth century. And, when this happened, the result came to be regarded as a watershed moment in the progress of rational thought. Such is the status of infinite sets in the mathematics of our time.

0.2 An analytic model of the straight line

It is a disposition of the human mind to understand things analytically. We take an object, separate it into what we think to be its principal elements, and then reconstruct it in terms of those elements. The question arises:

If we are to analyze the geometric line into its principal elements, what should the principal elements be?

The ancients had thought of the geometric line as an assemblage of infinitely many geometric points until it was realized by some, especially Zeno of Elea (5 BC), that their conception of the infinite was too naïve to make the notion suitable for building a non-contradictory analytic model of the line. In other words, constructing an analytic model of the line which captures its continuity while representing it as an assemblage of infinitely many geometric points was a task too intricate for the intellect of the times to handle. Thus infinite sets as conceptual tools of model building acquired a controversial status, which persisted for about 24 centuries.

It was not until the turn of the twentieth century that a positive attitude towards infinite sets began to be adopted by the vast majority of mathematicians. This followed a series of important developments in the nineteenth century. In 1888, using the already known device of one-to-one correspondence, Richard Dedekind (1831–1916) gave a precise definition of an infinite set. To wit, he called a set A infinite if it has a proper subset B that can be put into one-to-one correspondence with A. Then Georg Cantor (1845–1918) developed an abstract theory of sets in which the groundwork was laid for widespread acceptance of infinite sets as legitimate mental constructs that can serve as building blocks of various mathematical concepts. Cantor's development was later given a rigorous axiomatic treatment, in what is now called Zermelo–Fraenkel *set theory*, with the axiom of choice. This is a rich conceptual structure within which we can develop virtually the entire known mathematics. In particular, an analytic model of the straight line can be constructed within this structure. This model, which represents the line as an assemblage of infinitely many points while capturing its continuity, is, of course, none other than what we call today the *standard theory of the real numbers* and denote $(\boldsymbol{R}, +, \cdot, <)$. This system was developed in the second half of the nineteenth century independently by Karl Weierstrass (1815–1897), Georg Cantor, and Richard Dedekind. Their methods were different but produced equivalent structures, all using infinite sets as their building blocks.

Analytic models of the straight line, such as the familiar real number system, provide the foundation of *mathematical analysis,* which is in part the result of attempts to provide analytic theories of such basic intuitive notions as continuity and smoothness. The physical world around us may contain no straight lines, continuous curves, or smooth surfaces. But that does not matter. Our mathematical theories – our mental linguistic constructs –, which have risen from intuitive concepts such as these, have served us remarkably well as conceptual frameworks within which useful models of various aspects of the physical world can be (and have been) built. This point is much in evidence in the myriad applications of mathematical analysis. For example, we use the system $(\boldsymbol{R}, +, \cdot, <)$ to model

Newtonian time, the system $(\boldsymbol{R}^3, +, \cdot)$ to model Newtonian space, and the system $(\boldsymbol{R}^4, +, \cdot)$ to model Minkowski spacetime, which provides for the representation of physical entities conceived within Einstein's general theory of relativity. Indeed, today the student of classical mechanics needs to master the subject of analysis on finite-dimensional vector spaces, the student of quantum mechanics needs to master the subject of analysis on infinite-dimensional vector spaces, and the student of economic theory needs to master both.

0.3 The rise and fall of infinitesimals

In the sixteenth and seventeenth centuries there was an irrepressible urge among mathematicians to apply analytic methods in the study of continuous quantities. Thinking of a line or a curve as an assemblage of infinitely many points was still not an idea they could employ with confidence. Such analytic models faced objections such as these:

> How can a continuous object (such as a line segment), which has a nonzero magnitude, be reconstructed from points, which have zero magnitude? How can a geometric object (such as a plane) be assumed without contradiction to have been constructed from objects of lower dimensions (such as lines or points)?

Unable to overcome such hurdles, the mathematicians of the seventeenth century appealed to alternatives such as thinking of a continuous object as composed of an infinity of infinitesimal parts (the parts being of the same kind as the whole) or thinking of, for example, a curve as generated by the continuous motion of a point. Johannes Kepler (1571–1630) regarded a circular region as an infinity of infinitesimal triangles with a common vertex at the center and a spherical region as an infinity of pyramids with a common vertex at the center. Leibniz (1646–1716), one inventor of calculus, used the term *differential* and the corresponding symbol dx to refer to an infinitely small change in x. Using this notion he defined the *derivative* as the quotient $\frac{dy}{dx}$ and the *integral* as the sum of infinitely many infinitesimals ydx. These were, in turn, applied to solve the geometric problems of determining tangents, areas, and volumes. Newton (1642–1727) referred to a (continuous) variable x as a *fluent*, thinking of it as a quantity that is in continuous flow. He referred to its (instantaneous) rate of change as a *fluxion* and used the symbol $\dot{x}$ to denote it. He also used the symbol "o" to denote an infinitely small quantity of time. This led him to the notation $\dot{x}o$ for an infinitesimal change in x. Using these ideas, Newton was able to develop calculus (independently of Leibniz) and use it to account for Kepler's laws describing the motions of the planets.

The method of infinitesimals was a boon to its practitioners. It appealed to their intuition, simplified their calculations, and brought resolution to many hitherto inaccessible problems. However, just as the ancient Greeks had been unable to imbed their infinite sets in a theory that was free from the known paradoxes, so the mathematicians of the seventeenth and eighteenth centuries were unable to give a rigorous treatment of their infinitesimals. This led, in the latter part of the nineteenth century, to a revolution in the foundations of calculus, in the course of which infinitesimals were banished from mathematics for want of a sound logical standing. By the leadership of Weierstrass, the term *limit* defined in terms of $\epsilon\delta$ statements replaced the term "infinitesimal" in the definition of concepts as well as in arguments; and, as mentioned earlier, a rigorous theory of the real numbers (devoid of the terms *infinitely small* and *infinitely large*) was developed by Cantor, Dedekind, and Weierstrass. Thus, since the latter part of the nineteenth century the $\epsilon\delta$ method has become the *standard* method of defining concepts and proving theorems in analysis and the Cantor–Dedekind–Weierstrass number system $(\boldsymbol{R}, +, \cdot, <)$ has become the *standard* analytic model of the straight line.

0.4 The return of infinitesimals

Fortunately, it did not take too long for the ban on infintesimals to end. The intuitive appeal and simplicity of this notion kept it alive in the minds of some mathematicians during the twentieth century. The attempts to reinstate infinitesimals as legitimate concepts came to a definitive conclusion in the discovery of *nonstandard analysis* by Abraham Robinson (1918–1974). Robinson's nonstandard analysis, which is the result of a sophisticated application of the concepts and methods of contemporary mathematical logic, was introduced in 1960–1961 when Robinson was a professor of mathematics at the University of California, Los Angeles, and the Institute of Advanced Studies at Princeton. Robinson's first book on the subject published by North-Holland, appeared in 1966 under the title *Non-Standard Analysis*.

Nonstandard analysis (NSA) provides a rich framework for the mathematics of the infinite. The analytic models of continuous quantities that can be constructed within this framework capture at once both the intuition of the ancient Greek mathematicians and that of the mathematicians of the seventeenth century. It provides for the construction of a *new* real number system $({}^*\boldsymbol{R}, +, \cdot, <)$ that not only represents the line as an assemblage of infinitely many points (as the Greeks had done) but also enables us to visualize the line as containing points that are infinitely close and also points that are infinitely far apart (which accommodates what Leibniz, Newton, and their followers had ideated).[2] The system $({}^*\boldsymbol{R}, +, \cdot, <)$,

[2] It is of historical interest to note that Leibniz, Newton, and their followers thought of a continuous quantity, say, a line segment, as something that could be partitioned into an infinity of smaller line segments each having an infinitesimal length; while this idea can be accommodated within Robinson's analytic model of the line, it is not the same as the latter. In fact, they tried to avoid thinking of a continuous quantity as an assemblage of infinitely many points, for they did not know how to handle the paradoxes associated with such conceptions. The notion of a non-denumerable infinite set (which is essential for such analytic models of continuous quantities) was not available then.

Figure 0.1 Robinson (left) and Luxemburg at the 1970 conference on NSA at Oberwolfach (Photographs courtesy of Ron Luxemburg, photography by Pasadena, Los Angeles, http://rlux.com.).

as the new foundation of mathematical analysis, is preferable to the standard system because it allows us to study the continuity, smoothness, and measurement of areas and volumes (and numerous other concepts obtained by generalizing these) through theories that appeal to our intuition and have simpler syntactical structures.

The system $({}^*\boldsymbol{R}, +, \cdot, <)$ is a linearly ordered field that contains infinitely small and infinitely large numbers as well as an isomorphic copy of the standard system of the real numbers, $(\boldsymbol{R}, +, \cdot, <)$. The first such system was discovered by Edwin Hewitt (1920–1999) while he was studying residue class fields of continuous functions modulo free maximal ideals.[3]

Today we use an equivalent technique based on what is called a free ultrafilter. This technique was first used by W. A. J. Luxemburg (one of the co-founders of nonstandard analysis) in his early (1960–1962) lectures on NSA at California Institute of Technology; see Figure 0.1. This method of constructing the system $({}^*\boldsymbol{R}, +, \cdot, <)$ is discussed briefly in Section 0.5 below. As mentioned earlier, Robinson's original presentation of NSA was based on certain ideas from the field of mathematical logic that are not very familiar to most mathematicians. Luxemburg adapted nonstandard analysis to suit the modes of thinking of a wider community of mathematicians. This, and his many other contributions to NSA,

[3] Edwin Hewitt, Rings of real-valued continuous functions, *Trans. Amer. Math. Soc.* **64** (1948), 54–99.

expedited the spread and further development of the methods of modern infinitesimals.[4] Among Luxemburg's widely quoted works on NSA, one might mention his article A general theory of monads, in *Applications of Model Theory to Algebra, Analysis, and Probability*, edited by Holt, Reinhart and Winston, 1969.

0.5 Ultrafilters and ultrapowers

In the remaining of this introduction, we provide a brief overview of how ultrafilters may be used in constructing structures that allow a rigorous discussion of modern infinitesimals.

0.5.1 Definition (Ultrafilters)[5] A family $\mathcal{U}$ of subsets of the positive integers $\boldsymbol{Z}^+$ is called a *free ultrafilter* on $\boldsymbol{Z}^+$ if it satisfies the following conditions:

1. $\emptyset \notin \mathcal{U}$;
2. if $A, B \in \mathcal{U}$ then $A \cap B \in \mathcal{U}$;
3. if $A \in \mathcal{U}$ and B is a subset of $\boldsymbol{Z}^+$ that contains A then $B \in \mathcal{U}$;
4. if $S \subseteq \boldsymbol{Z}^+$ then $S \in \mathcal{U}$ or $S' = \{x \in \boldsymbol{Z}^+ : x \notin S\} \in \mathcal{U}$;
5. no finite subset of $\boldsymbol{Z}^+$ belongs to $\mathcal{U}$.

Now let $\boldsymbol{X}$ be any infinite set, and let $\boldsymbol{X}^{\boldsymbol{Z}^+}$ be the set of all the sequences (a_n) in $\boldsymbol{X}$. We define an equivalence relation $\equiv$ on $\boldsymbol{X}^{\boldsymbol{Z}^+}$ by writing $(a_n) \equiv (b_n)$ if and only if $\{n \in \boldsymbol{Z}^+ : a_n = b_n\} \in \mathcal{U}$. It is easy to see that $\equiv$ is indeed an equivalence relation. For example, to prove reflexivity (i.e., $(a_n) \equiv (a_n)$ for all $(a_n) \in \boldsymbol{X}^{\boldsymbol{Z}^+}$), we must show that $\{n \in \boldsymbol{Z}^+ : a_n = a_n\} \in \mathcal{U}$. This is an immediate consequence of conditions 1 and 4 of Definition 0.5.1 since the set $\{n \in \boldsymbol{Z}^+ : a_n = a_n\}$ is none other than the entire set $\boldsymbol{Z}^+$.

0.5.2 Definition The set $\mathbb{X}$ of all the equivalence classes of $\boldsymbol{X}^{\boldsymbol{Z}^+}$ that are induced by $\equiv$ is called an *ultrapower* of $\boldsymbol{X}$. For each $x \in \boldsymbol{X}$ the equivalence class $[(x, x, x, \ldots)]$ of the constant sequence $(x, x, x, \ldots)$ is denoted by ${}^*x$. An element $\boldsymbol{x} \in \mathbb{X}$ is called *standard* if there is an $x \in \boldsymbol{X}$ such that $\boldsymbol{x} = {}^*x$. The rest of the elements of $\mathbb{X}$ are called *nonstandard*. The collection of all the standard elements of $\mathbb{X}$ is denoted by ${}^\sigma\mathbb{X}$. Theorem 0.5.4 gives the condition that guarantees the existence of nonstandard elements in $\mathbb{X}$.

0.5.3 Remark The symbol ${}^*\boldsymbol{X}$ is an alternative notation for $\mathbb{X}$. We shall use this symbol particularly to denote an ultrapower ${}^*\boldsymbol{R}$ of the set of the ordinary real numbers $\boldsymbol{R}$.

0.5.4 Theorem *Let $\mathbb{X}$ be an ultrapower of the set $\boldsymbol{X}$. If $\boldsymbol{X}$ is infinite, then $\mathbb{X}$ has nonstandard elements.*

[4] For further remarks see *Abraham Robinson, the Creation of Nonstandard Analysis, a Personal and Mathematical Odyssey*, J. W. Dauben, Princeton University Press (1995), pp. 394–396.
[5] A proof of the existence of free ultrafilters requires the axiom of choice.

Proof Since $\boldsymbol{X}$ is infinite, there is a sequence (a_n) in it whose terms are distinct. Let $\boldsymbol{\alpha} = [(a_n)]$. We claim that $\boldsymbol{\alpha} \notin {}^{\sigma}\mathbb{X}$. That is, (a_n) does not belong to the equivalence class of any constant sequence $(x, x, x, \ldots)$. To see this, fix $x \in \boldsymbol{X}$. Since the set $S = \{n \in \boldsymbol{Z}^+ : x = a_n\}$ has at most one element by condition 5 of Definition 0.5.1, we have $S \notin \mathcal{U}$. This means that $(a_1, a_2, a_3, \ldots)$ and $(x, x, x, \ldots)$ are not related by $\equiv$. Hence $\boldsymbol{\alpha} \neq {}^*\boldsymbol{x}$. □

Now fix a free ultrafilter $\mathcal{U}$ on $\boldsymbol{Z}^+$, and let ${}^*\boldsymbol{R}$ denote the ultrapower of $\boldsymbol{R}$ that is obtained by means of $\mathcal{U}$. For convenience, for each sequence (a_n) in $\boldsymbol{R}$, let its equivalence class $[(a_n)]$ be denoted by the boldface symbol $\boldsymbol{a}$.

0.5.5 Definition Given $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in {}^*\boldsymbol{R}$, we write:

$\boldsymbol{a} = \boldsymbol{b}$	iff	$(a_n) \equiv (b_n)$;
$\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c}$	iff	$(a_n + b_n) \equiv (c_n)$;
$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{c}$	iff	$(a_n \cdot b_n) \equiv (c_n)$;
$\boldsymbol{a} < \boldsymbol{b}$	iff	$\{n \in \boldsymbol{Z}^+ : a_n < b_n\} \in \mathcal{U}$.

0.5.6 Hyperreal number system The system $({}^*\boldsymbol{R}, +, \cdot, <)$ that we have just defined is referred to as a *hyperreal number system*. It is a linearly ordered field, and contains an isomorphic copy of the system $(\boldsymbol{R}, +, \cdot, <)$.[6] This isomorphism assigns to each $a \in \boldsymbol{R}$ the equivalence class $\boldsymbol{a} = [(a, a, a, \ldots)]$. The set ${}^*\boldsymbol{R}$ has *unlimited elements* (or *elements with infinitely large magnitudes*). Such elements have absolute values that are greater than every positive element of ${}^{\sigma}[{}^*\boldsymbol{R}]$. For example, if $\omega = [(1, 2, 3, \ldots)]$ then we have $\omega > \boldsymbol{a}$ for all $\boldsymbol{a} \in {}^{\sigma}[{}^*\boldsymbol{R}]$ since $\{n \in \boldsymbol{Z}^+ : n > a\} \in \mathcal{U}$ for each $a \in \boldsymbol{R}$. The reciprocals of unlimited numbers are *infinitesimals*. Thus, for example,

$$\frac{1}{\omega} = \left[\left(1, \frac{1}{2}, \frac{1}{3}, \ldots\right)\right]$$

is an infinitesimal – it is smaller than every standard positive number in ${}^*\boldsymbol{R}$. Two numbers $\boldsymbol{a}, \boldsymbol{b}$ in $\mathbb{R}$ are infinitely close, which is written as $\boldsymbol{a} \simeq \boldsymbol{b}$, if $|\boldsymbol{a} - \boldsymbol{b}|$ is an infinitesimal. Our new real number system allows us to visualize the geometric line as in Figure 0.2.

0.6 What is internal set theory?

To do mathematical analysis with the help of infinitesimals, we need much more than just an ultrapower of $\boldsymbol{R}$. Indeed, a framework that turns out to be adequate for

[6] A proof of this statement can be found in *An Introduction to Nonstandard Analysis* (p. 5), by Albert E. Hurd and Peter A. Loeb, Academic Press, 1985.

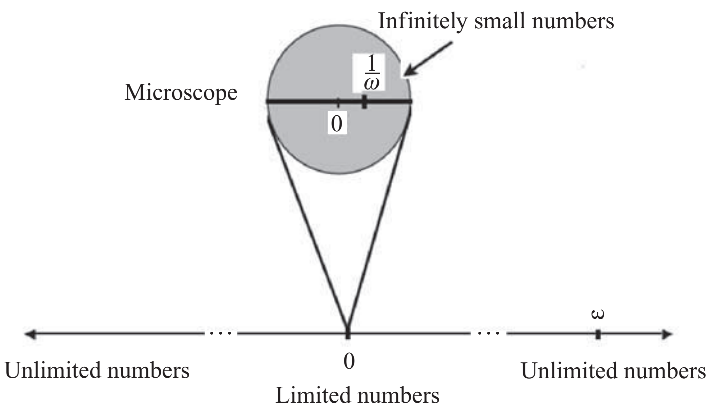

Figure 0.2 Infinitesimals viewed through a microscope.

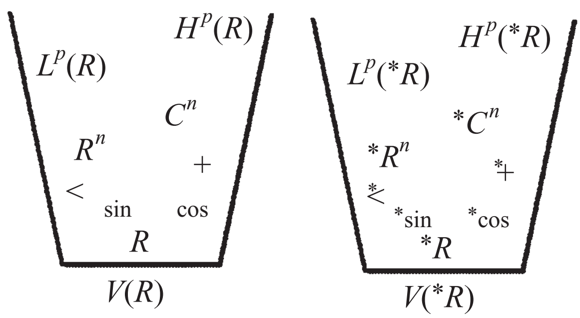

Figure 0.3 Superstructures over $\boldsymbol{R}$ and ${}^*\boldsymbol{R}$.

virtually every mathematical construction is the result of a sophisticated ultrapower construction involving *superstructures,* which are defined as follows.

0.6.1 Definition Given an infinite set X, the *superstructure* $V(X)$ on X is the set $V(X) = \bigcup_{n=0}^{\infty} V_n$, where (V_n) is defined recursively as:

$$V_0 = X, \qquad V_n = V_{n-1} \cup \mathcal{P}(V_{n-1}), n \geq 1.$$

Here $\mathcal{P}(V_{n-1})$ is the set of all the subsets of V_{n-1}.

The set $V(\boldsymbol{R})$ contains a representation of virtually any object discussed in the traditional treatments of mathematical analysis; furthermore, virtually any construct that provides for an application of modern infinitesimals in analysis has a representation within $V({}^*\boldsymbol{R})$. We may picture this as in Figure 0.3.

To employ the methods of modern infinitesimals in analysis, one needs to be well versed not only in the properties of such ultrapowers as ${}^*\boldsymbol{R}$, ${}^*\boldsymbol{R}^n$, and $L^p({}^*\boldsymbol{R}^n)$ but also in the interplay between the two structures $V(\boldsymbol{R})$ and $V({}^*\boldsymbol{R})$. The theory that

Figure 0.4 Edward Nelson.

describes[7] this interplay employs sophisticated ideas from model theory (a branch of mathematical logic) and might be difficult for someone without a background in that subject. This situation has motivated attempts to find approaches to NSA that are even simpler than that established by Luxemburg.

Among these attempts, the most successful has been Edward Nelson's axiomatic approach, which is called *internal set theory* (IST). Edward Nelson (Figure 0.4) is a contemporary mathematician currently working at Princeton University. His original account of IST was published under the title Internal set theory, a new approach to nonstandard analysis, in the *Bulletin of the American Mathematical Society* **83** (1977), 1165–1198. He showed, in this article, that a model of IST can be constructed within ZFC. Internal set theory is quite easy to master, as the reader will soon see.

To gain a sense of what IST describes, we need to know a little more about what is going on inside $V({}^*\boldsymbol{R})$. This is discussed next.

0.7 Internal, external, and standard sets

In this section, we provide an overview of the kind of structure that internal set theory describes axiomatically. Let $\boldsymbol{X} = V_0 \cup V_1 \cup V_2$, where

$$V_0 = \boldsymbol{R}, \quad V_1 = V_0 \cup \mathcal{P}(V_0), \quad \text{and} \quad V_2 = V_1 \cup \mathcal{P}(V_1).$$

As examples of elements of $\boldsymbol{X}$, one could mention the number $a = 2$, the open interval $A = (0, 3)$, and the family $\mathcal{A} = \mathcal{P}((0, 3))$ of all the subsets of the interval $(0, 3)$. Notice that $a \in A \in \mathcal{A}$.

Now fix a free ultrafilter $\mathcal{U}$ on $\boldsymbol{Z}^+$, and let $\mathbb{X}$ be the ultrapower of $\boldsymbol{X}$ that corresponds to $\mathcal{U}$.

[7] See p. 70 of *An Introduction to Nonstandard Analysis*, by Albert E. Hurd and Peter A. Loeb, Academic Press, 1985.