SCIENCE HISTORY PODCAST: *NUMBERS AND NUMBER SYSTEMS*

1. Introduction

In this podcast episode, we will discuss the history and development of Number Theory, viewed through the lens of numbers and number systems. We will begin with the *natural numbers*, originally

$$\mathbf{N} := \{1, 2, 3, \ldots\}$$

which arose to abstractly represent and manipulate tallies. Over the course of thousands of years, this system of counting numbers was studied, expanded, and generalized to include zero, negative numbers, rationals, real numbers, irrational, algebraic and transcendental numbers, complex numbers, modular arithmetic, and p-adic numbers. All of these systems of numbers play an important role in *Number Theory* which, at its core, is the study of $\mathbf{N}$ and its properties. The goals of this episode are:

(1) Provide a mathematical and historical overview of these different number systems and their development.
(2) Highlight the challenges and obstacles that mathematicians and civilizations faced with new concepts of *number*.
(3) Give the listener a sense of why these various systems of numbers are interesting and important.
(4) Touch on some of the important unsolved problems in modern number theory, and how these different number systems play a role.

A good overview reference for some of the history is Wikipedia. Hopefully the right-margin of this document is also a decent reference.

2. The integers

Actually, there is some debate about the definition of $\mathbf{N}$. Many texts define $\mathbf{N}$ as above, but I would argue that the "right" definition should include zero, as below.

$$\mathbf{N} := \{0, 1, 2, 3, \ldots\}$$

There are a few good arguments for this: the natural numbers are the "counting" numbers, that is, they are exactly the numbers which are cardinalities of finite sets (modern mathematics is predicated on axiomatic set theory, typically using the ZF or ZFC axioms). In set theory, one of the axioms is that the intersection of two sets should always be a set, and that necessitates having the *empty set*, $\emptyset$, which contains no elements and has cardinality 0. So if $\mathbf{N}$ is the set of cardinalities of finite sets, it must include zero.

Another argument has to do with arithmetic: the natural numbers are closed under addition and multiplication, and the element 1 is a *multiplicative identity* as it satisfies $1 \cdot n = n \cdot 1 = n$ for every natural n. From that

Alternatively denoted $\mathbb{N}$. Note that whenever we write something like $A := B$, we mean that the definition of A is B. Sometimes this is written $A \triangleq B$.

There are two kinds of 0: the number representing *nothing*, as in $1 - 1 = 0$, and a *placeholder* for multiplication, as in the decimal $1206 = ((12)^{*}\ (100)) + 6$. The earliest known placeholder zero is on a Babylonian tablet from $\approx$ 400BC, that uses a base-60 number system with $''$ for a placeholder. The Greeks mostly ignored 0 because they used geometric lengths (e.g. of string) rather than symbolic numbers. Ptolemy used a placeholder 0 in 130AD, but only as punctuation. The Mayans discovered the placeholder 0 by 665AD, while the Indians began using it by 876AD. By 830AD the Indians had discovered the number 0, but they did not understand the impossibility of division-by-0 until more than 500 years later. Fibonacci, an Indian mathematician, used both kinds of 0, but called 0 a "sign" rather than a number. Al-Khwarizmi, the namesake of "algorithm" and "algebra", brought the Hindu 0 to Iraq in the 12$^{\text{th}}$ century. Qin Jiushao brought the Hindu 0 to China in 1247. [OR00]

A *set* is an unordered list of symbols with no duplicates, which does not contain itself. So, the infamous "set of all sets" is not a set at all, as it would have to contain itself. The Russell-Zermelo paradox (discovered around 1903) considered the set of all sets that do not contain themselves. This strange set is apparently a member of itself, if and only if it is not a member of itself. ZFC is a set of axioms under which the paradox disappears, because the pathological set (and others like it) need not exist in the first place. [ID20]

Date: April 25, 2021.

point of view, it makes sense that $\mathbf{N}$ should also contain an *additive identity*, which satisfies $x + n = n + x = n$ for all naturals n. Of course, $x = 0$ is the unique value that works!

This second argument will return as a central, key theme:

Philosophy 2.1. Mathematics often isolates some property that we would like a number to satisfy, and if we don't already have any number to satisfy it, we *make* a new number and larger number system to verify the property.

Such is the genesis of the integers

$$\mathbf{Z} := \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\},$$

consisting of the naturals (with zero!) and their negatives. It took civilization some time to accept negative quantities as "numbers" in their own right, but the above philosophy plays a critical role here. Addition is fundamental operations of the number system $\mathbf{N}$, so questions like "is there a number x for which $x + 3 = 5$" arise naturally. Of course, in this case $x = 2$ is the unique solution. But when we reverse the roles of 3 and 5, as in the equation $x + 5 = 3$, suddenly there is *no solution* if we only allow ourselves to work in the natural numbers! So a solution is *created*, and called -2, and it satisfies the *defining* attribute that $2 + (-2) = 0$.

I don't think that negative integers *exist* in the same way that positive integers do: you can *show* me 1 apple, you can hold it in your hand, you can give it to me and I can take a bite out of it. But you can't do any of that with -1 apples, and one has to invoke notions of debt (already a most abstract notion!) to make "real world" sense of negative numbers. One of the central points I'd like to make is that mathematics is *best*—most complete, powerful, useful, beautiful—when we don't worry so much about "real world" meaning or application of mathematical ideas or theories. That is a counterintuitive idea, as the real world has informed and spurned so much important mathematics, but the mathematics itself develops best untethered from these shackles of its origins.

Now the integers are a lot better than the naturals, because they form a *ring*. A ring is a set of numbers in which one can do arithmetic (addition and multiplication) *and* there are both additive and multiplicative identities, *and* every element has an additive inverse, meaning that we can always solve the equation $x + r = 0$ for any number r. One also asks that the usual axioms (associativity, distributivity of multiplication over addition etc.) hold. For the purposes of this discussion, we'll only talk about *commutative* rings, *i.e.* rings in which

$$x \cdot y = y \cdot x$$

for all x and y. There are lots of important examples of non-commutative rings, like rings of square matrices, but that is a topic for another time.

3. The rationals

Being a ring, the integers are closed under addition and multiplication, and every integer has an additive inverse, but very fee integers have *multiplicative* inverses. Said differently, we can solve equations like $4x = 12$ because 12 is an integer multiple of 4, but change that 12 to a 10 and we are out of

luck! So it is reasonable to create a new number system in which every equation $mx = n$ with integers m and n has a solution. Actually, there is no such number system, because having a solution to $0 \cdot x = 1$ causes big problems, and many of the desirable properties of a "number system" break down. For example, if x satisfies the above equation, so does $x + n$ for any integer n, so either there are infinitely many x's that solve the equation, or we have to accept that $x = x + n$ for all n, and either we aren't allowed to subtract x from both sides (ruining basic arithmetic), or in our new system of numbers $n = 0$ for all n, and that is not very useful! So by *definition*, the rationals are all solutions x to equations $mx = n$ with $m \neq 0$. We represent such a solution as the fraction n/m, but then we have some weird rules to maintain the usual rules of arithmetic, since if $mx = n$ then $2mx = 2n$, so $2n/2m = n/m$.

$$\mathbf{Q} := \left\{ \frac{n}{m} \ : \ n, m \in \mathbf{Z}, m \neq 0 \right\} /$$

It took a long time to really understand fractions, because strictly speaking a *fraction* like $1/2$ is an equivalence class of pairs of integers

$$\frac{1}{2} = \{(n, m) \ : \ n, m \in \mathbf{Z}, m = 2n\} = \{(1, 2), (2, 4), (3, 6), (-1, -2), (-5, -10), \ldots\}$$

This is just saying that the fractions $1/2, 2/4, 3/6, -1/-2, \ldots$ are all the *same* number, because they solve mathematically equivalent equations. But they don't teach equivalence classes in school when we learn about fractions, and I think this is a difficult and subtle concept.

The set of rational numbers forms what is called a *field*: it is a ring in which every nonzero number has a multiplicative inverse. Fields have the satisfying property that for any elements a, b, c, the linear equation $ax+b = c$ always has a solution, at least as long as $a \neq 0$.

The advent of the decimal system allows us to represent rationals as decimals, like $1/2 = 0.5$ and $1/3 = 0.\overline{3}$, and there is a rather nice characterization of which decimal expansions are rational numbers:

Proposition 3.1. *A decimal expansion comes from a rational number if and only if it either terminates, or is eventually repeating.*

One confounding feature of decimals is that, just like a rational number can have many representations as a fraction ($1/2 = 2/4 = 3/6$), so too a number can have more than representation as a decimal expansion! A classic example is:

$$1.\overline{0} = 0.\overline{9}.$$

This example is really frustrating for many people, and has resulted in more than one shouting match and even tears. For whatever reason, many people *believe* that every number has a unique decimal expansion, possibly because we are taught from an early age to equate the concepts of *number* and *decimal expansion*. But it just isn't true, much in the same way that rational numbers don't have a unique representation as a fraction. In grade-school, one of my math teachers *insisted* that 1 and $0.\overline{9}$ are *different* numbers, and no amount of reasoning would convince her otherwise. But if you admit that decimal expansions represent numbers, and therefore must obey the basic laws of

arithmetic (these are axiomatized!), then $1 = 0.\overline{9}$ is forced upon you, since:

$$\frac{1}{3} = 0.\overline{3}$$

by long division, so multiplying both sides by 3 yields $1 = 0.\overline{9}$. Said another way, whatever number $x = 0.\overline{9}$ is, it has to obey the rules of arithmetic, so has to satisfy $10x - x = 9$ since multiplication by 10 shifts the decimal to the right by 1, and subtracting x from this lops off the infinite repeating tail of $9's$. But then $9x = 9$, which again forces $x = 1$.

4. Real numbers

With Proposition 3.1 in mind, it is natural to ask what numbers are represented by *arbitrary* decimal expansions. These are of course the real numbers: $\mathbf{R}$, and they solve a basic problem with the rationals, which is that the rational numbers have *holes*. That is, there are infinite sequences of rational numbers which are getting arbitrarily close to *something*, but that something isn't a rational number! For example:

$$3, 3.1, 3.14, 3.141, 3.1415, \cdots$$

This just reflects the fact that if you truncate an arbitrary decimal expansion at any point, you will get a terminating decimal expansion, which must therefore be a rational number. From this point of view, the rational numbers are *not complete* with respect to the usual measure of distance given by absolute value:

$$d(x, y) := |x - y|.$$

The reason is that there are lots of sequences of rationals as above, with the property that the distance between any two terms in the sequence gets smaller and smaller as you go out, but there is no "final term", or limit, of the sequence within the system of rational numbers. In this way, the real numbers are the *completion* of the rationals with respect to the usual distance function above: they "fill in" all of the holes that the usual distance function perceives in the form of sequences whose terms get increasingly close together. Such sequences are called *Cauchy sequences*, after the French mathematician Augustin-Louis Cauchy (1789–1857), who was a pioneer of mathematical rigor in algebra and analysis, and introduced the much-loathed (by students) notion of $\epsilon - -\delta$ arguments to Calculus.

The system of real numbers has many wonderful properties, including *continuity*, and therefore provides the right setting in which to do calculus. But they have some rather bizarre and frustrating features: there are continuous functions that are differentiable nowhere, nonzero smooth functions which vanish identically on any closed subset, nonnegative integrable functions with zero integral, and so forth.

```
Discussion of irrational, algebraic and transcendental
numbers?
```

ToDo

5. p-Adic numbers

As explained above, the real numbers are created by "filling in the holes" of the rational numbers that are detected by the usual distance function $d(x, y) = |x - y|$. But this is *not* the only distance function on the rational numbers, and different distance functions will pick up on different "holes", which are filled in by entirely new systems of numbers! To motivate this, which will seem very strange at first given our rich experience of the real numbers, consider the rational number 0. This is arguably the *smallest* rational number by any sense of small since adding it to any other rational number does nothing. But there is another *arithmetical* measure of smallness that zero satisfies: it is the only integer that is *arbitrarily divisible*, in the sense that 0 is a multiple of n for every integer n. Thinking in terms of decimals, 0 is the only integer which, when represented in base 10, has arbitrarily many zeros to the left of the decimal point. Said in terms of arithmetic, 0 is the only integer which is divisible by 10^k for every natural number k. This inspires a new prototype of distance:

Definition 5.1. The 10-adic absolute value of an integer n is the $1/10^k$, where 10^k is the highest power of 10 dividing n.

By definition, the 10-adic absolute value of 0 is 0. This notion of size is really different from the usual one, since 1000 is smaller than 100, which is smaller than 10, and 311 and 7 have the same size! But actually, 10 is a bad choice, because we don't really get a well-behaved distance function from it. Indeed, the 10-adic absolute values of 2 and 5 are both 1, but their product has 10-adic absolute value 0.1, so this new absolute value isn't multiplicative. That causes all kinds of problems, and it doesn't really deserve to be called an absolute value. But this problem is caused by the fact that 10 can be factored as $2 \cdot 5$, so it completely disappears if we replace 10 by a prime number p. In this way, we get for every prime p, a p-adic absolute value on $\mathbf{Z}$, which can be extended to $\mathbf{Q}$ by declaring that the absolute value of a ratio is the difference of the absolute values of numerator and denominator:

$$|n|_p := 1/p^k \text{ where } p^k \text{ is the highest power of } p \text{ dividing } n$$
$$|m/n|_p := |m|_p - |n|_p$$

For example, using the 3-adic absolute value, the sequence $3, 9, 27, 81, 243, \ldots$ is approaching 0, while the sequence $2, 4, 8, 16, 32, \cdots$ is a sequence of numbers all of the same 3-adic size 1. So for each prime p, we get a new distance function on the rational numbers: $d_p(x, y) = |x - y|_p$, and these new distance functions "see" new holes in the rational numbers, which can be filled in (like we made the real numbers) to produce the *p-adic numbers* $\mathbf{Q}_p$, which is a completely new number system for each prime p. Said differently, $\mathbf{Q}_p$ is the completion of $\mathbf{Q}$ with respect to the p-adic absolute value. We now have infinitely many different ways of filling in the holes in the rationals:

$$\mathbf{R}, \mathbf{Q}_2, \mathbf{Q}_3, \mathbf{Q}_5, \mathbf{Q}_7, \mathbf{Q}_{11}, \ldots$$

and an amazing theorem of Alexander Ostrowski (1893–1986) asserts that these are the *only* completions of $\mathbf{Q}$; *i.e.* that there are no other ways to fill in the holes.

The number systems $\mathbf{Q}_p$ are really different from $\mathbf{R}$, largely because the p-adic distance function is *ultrametric* in the following sense. The usual absolute value, which completes the rationals to the reals, satisfies the *triangle inequality*: the length of the hypotenuse of a triangle is at most the sum of the lengths of the other two sides. But the p-adic absolute value satisfies the *strong triangle inequality*, in which the hypotenuse of a triangle has length at most the minimum of the other two sides! No sum needed! This leads to the remarkable fact that the collection of all p-adic numbers of p-adic absolute value at most 1 forms a ring, so is a number system in its own right:

$$\mathbf{Z}_p := \{x \in \mathbf{Q}_p \ : \ |x|_p \leq 1\}$$

In fact, the p-adic absolute value sees "holes" already in $\mathbf{Z}$, and $\mathbf{Z}_p$ is what you get when you fill in these holes. For example, the sequence

$$1, 4, 13, 175, 4549, 11110 \ldots$$

converges to a limit L in $\mathbf{Z}_3$ because the successive differences are $3, 3^2, 2 \cdot 3^4, 2 \cdot 3^7, 3^8 \ldots$, which have 3-adic absolute values tending to zero. Actually, the 3-adic limit of this sequence L satisfies $L^2 = 7$, so that $\sqrt{7}$ exists in $\mathbf{Z}_3$; this can be seen by looking at $a^2 - 7$ for each a in the sequence above, and we get the integers

$$-2 \cdot 3, 3^2, 2 \cdot 3^4, 2 \cdot 7 \cdot 3^7, 2 \cdot 19 \cdot 83 \cdot 3^8, \ldots$$

which are progressively more and more divisible by 3, so 3-adically are tending to zero.

6. What are the p-Adic numbers good for?

Arithmetic—addition and multiplication—lead (as we have seen) to algebraic equations like $nx = m$ and $ax + b = c$, and eventually to more complex *systems of algebraic equations* like

$$x^2 + y^2 = z^2$$

Broadly construed, (algebraic) number theory is the study of such systems and their solutions. Solving a system like tha above with x, y, z *real numbers* is easy! We can pick more or less any values for x and y that we like, and then just solve for z. This leads to lots of real solutions, like

$$(x, y, z) = (1, 1, \sqrt{2}), (1, 2, \sqrt{5}), (2, 3, \sqrt{13}), \cdots$$

but it is really hard (or even impossible with many systems) to find any (nontrivial[1]) *rational* or *integer solutions*. Of course, the system above is a really famous one, whose integer solutions are exactly the *Pythagorean triples*, like $(3, 4, 5)$ and $(5, 12, 13)$. It is important to emphasize that computers are not very helpful in finding such integer solutions: because there are infinitely many integers, we can't have a computer just plug in numbers and hope that it will eventually stumble upon an integer solution, since such a procedure might never finish, and even if there are integer solutions to be

[1]The obvious solutions $(\pm 1, 0, \pm 1)$, $(0, \pm 1, \pm 1)$, and $(0, 0, 0)$ are considered trivial because they are not particularly interesting!

found, they are very rare compared to the real solutions, so there isn't much chance to find them by happenstance.

It turns out that for any system of algebraic equations like the one above, finding solutions in $\mathbf{Q}_p$ for each prime p, or determining that none exist, is about as easy as finding solutions in real numbers. Roughly speaking, this is because $\mathbf{Q}_p$—like the real numbers—is complete with respect to an absolute value, so doesn't have any "holes", and we can more or less plug in randomly and then solve for the missing values. This doesn't always work, just as it doesn't always work in the reals (for example, $x^2 + y^2 + z^2 = -1$ has no real solutions at all). But finding p-adic solutions to systems of equations, or determining that none exist, is a tractable problem, just as it is in the reals (for example, one has a p-adic Newton's method for producing solutions from approximations by an iterative procedure).

This same problem for *integer* solutions is called *Hilbert's tenth problem*, and was formulated by the influential German mathematician David Hilbert in his now eponymous list of problems, presented to the International Congress of Mathematicians in 1900. In 1970, building on prior work of many mathematicians, Matiyasevich finally proved in 1970 that Hilbert's tenth problem is *unsolvable*, meaning that there can be no algorithm to decide if a general system of algebraic equations has an integer solution or not. Amazingly, Matiyasevich's proof uses Fibonacci numbers in an essential way!

Nonetheless, an amazing result of Hasse and Minkowski *does* provide an algorithm for equations of small degree:

Theorem 6.1. *If an equation in finitely many variables with rational coefficients involves only the variables and their squares, then it has a rational solution if and only if it has a solution in* $\mathbf{R}$ *and in every* $\mathbf{Q}_p$.

So, for example, the Pythagorean equation above has terms only of degree 2, so the fact that it has solutions in $\mathbf{R}$ and in every $\mathbf{Q}_p$ implies that it has rational (end even integer) solutions.

This idea—that we can solve algebraic equations in rational numbers if we can do so in every completion of the rationals— is called the *Hasse principle*, and is a central philosophy in number theory. Unfortunately, it doesn't always hold! For example, Selmer found the equation

$$3x^3 + 4y^3 + 5z^3 = 0$$

has nontrivial solutions in $\mathbf{R}$ and in every $\mathbf{Q}_p$, but no nontrivial rational solutions at all!

References

[ID20] Andrew David Irvine and Harry Deutsch. Russel's paradox. https://plato.stanford.edu/entries/russell-paradox/#RPCL, October 2020. Accessed 26 April 2021.

[OR00] J. J. O'Connor and E. F. Robertson. A history of Zero. https://mathshistory.st-andrews.ac.uk/HistTopics/Zero/, November 2000. Accessed 25 April 2021.