Notes on Set Theory, Logic, and Computation

Alexander A. Stepanov
Daniel E. Rose

August 29, 2013

Contents

Authors' Note

This document contains material from the Three Algorithmic Journeys course taught by Alex Stepanov at A9 in 2012, which we decided to remove from the forthcoming book based on the course.

1 Tackling the Infinite

> No one shall expel us from the Paradise that Cantor has created.
>
> — David Hilbert, *Über das Unendliche Mathematische Annalen* 95 (1925)

Peano used a tricky word "set" (*classe* in his language) when he wrote the axioms of arithmetic. Other mathematicians used the word before, but tended to avoid talking about infinite sets. Could you take the entire group of natural numbers and treat them as a single entity? This seemed like a dangerous idea. The Greeks had largely ignored infinities — everything had to be done in a finite number of steps. But infinite sets, and other types of infinities, kept cropping up in mathematics.

In the late 1340s and early 1350s, Europe was devastated by an outbreak of bubonic plague, a pandemic that killed 30-60% of the population. French and English kings battled over control of Normandy, Brittany, and other parts of France in the Hundred Years War. Popes of the Catholic Church ruled from Avignon rather than Rome, eventually leading to a schism where both popes and "anti-popes" competed for power. Despite this tragedy and turmoil, the 14th century was actually a time when the first stirrings of the Renaissance could be felt. In Italy, Dante, Bocaccio, and Petrarch were writing their great works. Great logicians such as William of Ockham and Jean Buridan were teaching. And in France, a bishop named Nicole Oresme began working on how to solve infinite series.[1]

Consider the series

$$\sum_{i=1}^{\infty} \frac{i}{2^{i-1}} = 1 + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{4} + \cdots + n \cdot \frac{1}{2^{n-1}} + \ldots$$

How could we compute the sum? Oresme devised the following strategy: First, he rear-

[1] Archimedes had dealt with infinite series, but only in the context of geometry. Oresme was the first to deal with them algebraically.

ranged and grouped the terms like this:

$$\begin{aligned}\left(1+\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \left(\frac{1}{2}+\frac{1}{4}+\cdots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \left(\frac{1}{4}+\cdots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \vdots\end{aligned}$$

Notice that by adding columns vertically, there is 1 copy of 1, then 2 copies of $\frac{1}{2}$, then 3 copies of $\frac{1}{4}$, and so on, just as the original expression specifies.

Then he notes that each row is $\frac{1}{2}$ of the row above, so we can rewrite it like this:

$$\begin{aligned}1\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \frac{1}{2}\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \frac{1}{4}\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\dots+\frac{1}{2^{n-1}}+\dots\right)&+\\ \vdots\end{aligned}$$

Since every coefficient is applied to the same quantity, we can use the distributive law to pull out all the coefficients and put them together into a sum, like this:

$$\left(1+\frac{1}{2}+\frac{1}{4}+\dots\right)\cdot\left(1+\frac{1}{2}+\frac{1}{4}+\dots\right)=$$

Notice that the first term, made up of all the coefficients, is the same as the second term; they're both $1+\frac{1}{2}+\frac{1}{4}+\dots$. But this sum was already common known to be equal to 2. So our original sum is $2\cdot 2$, or 4.

Nicole Oresme (1320-1382)

As we saw above, the 14th century was a period of warfare, death, and political and religious upheaval in France. Despite these inauspicious circumstances, a French priest named Nicole Oresme (pronounced "oh-RHEM") made major contributions to mathematics, physics, and economics during this period.

Oresme studied at one of the world's first universities, the Sorbonne in Paris, which was the intellectual center of Europe. He was influenced by the philosopher Jean Buridan. At the time, Buridan and a few like-minded scholars at Merton College in Oxford expanded their scope beyond studying theology and commenting on Aristotle, and began working on what we'd now call physics. Buridan challenged Aristotle's notion that a force must be constantly applied to keep a body in motion; he developed an early idea of inertia, which eventually led to Newton's first law of motion.

Oresme proved the *mean speed theorem*, also called the *Merton theorem*, which determines the average velocity of an object under constant acceleration. Oresme was the first to represent these quantities graphically, for example showing how the area under the velocity curve expressed distance.

After some years teaching at the Sorbonne, Oresme was appointed chaplain to the heir of the French throne, and continued in this appointment after the dauphin was crowned King Charles V. At the king's request, Oresme translated several of Aristotle's major works from Latin into French. This was the first time since ancient times that anyone had made important scientific literature available in the language of ordinary people. Oresme's work also helped French become an important scientific language for centuries to come.

Oresme also studied astronomy, and wrote a book in which he demonstrated that the Earth's rotation on its axis was as logically and practically sound an idea as the rotation of celestial spheres. Oresme's results laid the groundwork for Copernicus 100 years later. Oresme also wrote an important critique of Astrology, showing its logical flaws.

Later in life, Oresme became interested in Economics, and wrote a book describing what we would now call monetarism. He argued that kings (governments) shouldn't use monetary policy (minting more money) to get revenue; if they want revenue they should collect taxes, not debase the currency.

Today, Oresme is regarded as one of the founders of modern science.

Oresme was also the first to demonstrate the divergence of the Harmonic series $\sum_{i=1}^{2^n} \frac{1}{i}$.

Again, his approach involved re-grouping the terms:

$$\begin{aligned}\sum_{i=1}^{2^n} \frac{1}{i} = 1 + \frac{1}{2} & \\ & + \left(\frac{1}{3} + \frac{1}{4}\right) \\ & + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) \\ & + \dots \\ & + \left(\frac{1}{2^{n-1}+1} + \dots + \frac{1}{2^n}\right) \geq \frac{n}{2}\end{aligned}$$

Oresme observed that since each group of terms is greater than $\frac{1}{2}$, the sum is greater than $\frac{n}{2}$, and we know that as n grows, the number of groups grows. Therefore the series diverges.

Oresme's work on infinite series led to mathematical ruminations about infinity. He was the first to introduce the idea on which all modern set theory is based: the idea of a *one-to-one correspondence* to measure the cardinality[2] of infinite sets. Oresme observed that you can count odd integers with integers, so there are in some sense the same number of odd integers as integers. This was a controversial idea, since Euclid taught that the whole is greater than any of the parts, and odd integers are in some sense only a part of integers.

2 Paradoxes of the Infinite

> This is one of the difficulties which arise when we attempt, with our finite minds, to discuss the infinite, assigning to it those properties which we give to the finite and limited; but this I think is wrong, for we cannot speak of infinite quantities as being the one greater or less than or equal to another.
>
> – Galileo

About three hundred years after Oresme, the great Italian scientist Galileo wrote two revolutionary books. In one, *Dialogue Concerning the Two Chief World Systems*, he promoted the Copernican view that Earth and other heavenly bodies revolve around the Sun.[3] In the other, *Discourses and Mathematical Demonstrations Relating to Two New Sciences*, he discussed the strength of materials and the motion of objects, marking the beginning of modern Physics.

[2] The cardinality of a set is a measure of the number of elements it contains.

[3] Interestingly, he insisted that the orbits are circular, while Johannes Kepler, writing about the same time, realized they are elliptical.

In the *Two New Sciences* book, his final masterpiece, Galileo observed that there is a one-to-one correspondence between natural numbers and squares:

$$n \longleftrightarrow n^2$$
$$\{1, 2, 3, \dots\} \longleftrightarrow \{1, 4, 9, \dots\}$$

Yet the density of squares goes down as we go up. As Galileo put it:

> The proportionate number of squares diminishes as we pass to larger numbers, Thus up to 100 we have 10 squares, that is, the squares constitute 1/10 part of all the numbers; up to 10000, we find only 1/100 part to be squares; and up to a million only 1/1000 part; on the other hand in an infinite number, if one could conceive of such a thing, he would be forced to admit that there are as many squares as there are numbers taken all together.

If there are fewer and fewer squares as we count up, and the number approaches zero as we go to infinity, how can there be as many squares as numbers? This is one of many paradoxes of infinity that we'll encounter, and it illustrates that our intuitions about infinities are somehow broken.

A related paradox was described by Littlewood hundreds of years later in his book *A Mathematician's Miscellany*:

> (5) *An infinity paradox.* Balls numbered 1, 2, ... (or for a mathematician the numbers themselves) are put into a box as follows. At 1 minute to noon the numbers 1 to 10 are put in, and the number 1 is taken out. At $\frac{1}{2}$ minute to noon numbers 11 to 20 are put in and the number 2 is taken out. At $\frac{1}{3}$ minute 21 to 30 in and 3 out; and so on. How many are in the box at noon?

Intuitively, it seems that since at every step we put in more balls than we take out, there should be an infinite number of balls at the end. But at the same time, at the kth step we take out the ball labeled k, so for every ball there is a time it gets taken out. Therefore every ball is eventually taken out, and at the end we will have zero balls in the box.

3 Levels of Infinity

While other mathematicians deliberately avoided thinking about infinity, in a book called *Paradoxes of the Infinite*, Czech mathematician Bernard Bolzano exploring the idea of infinite sets in the early 19th century. He was also the first to use the term "one-to-one correspondence." Here are a few of his propositions:

> **§19** Not all infinite sets are equal with respect to their multiplicity.
>
> One could say that all infinite sets are infinite and thus one cannot compare them, but most people will agree that an interval in the real line is certainly a part and thus agree to a comparison of infinite sets.

§20 There are distinct infinite sets between which there is 1-1 correspondence. It is possible to have a 1-1 correspondence between an infinite set and a proper subset of it.

$y = \frac{12}{5}x$ and $y = \frac{5}{12}x$ gives a 1-1 correspondence between [0,5] and [0,12].

§21 If two sets A and B are infinite, one can not conclude anything about the equality of the sets even if there is a 1-1 correspondence.

If A and B are finite and A is a subset of B such that there is a 1-1 correspondence, then indeed $A = B$.

The above property is thus characteristic of infinite sets.

The last point, §21, is often known as "Dedekind's Axiom," but it came from Bolzano. Basically, he realized that the property of being able to have a 1-1 correspondence with its proper subset is what *defines* an infinite set.

Bernard Bolzano (1781–1848)

Bernard Bolzano was a remarkable mathematician, philosopher, theologian, and social reformer. He was born in Prague, in what is now the Czech Republic. At the time, Prague was in the Kingdom of Bohemia, which was considered a relatively unimportant part of the powerful Austro-Hungarian Empire.

Bolzano studied both mathematics and religion, ultimately deciding to become a priest. He was ordained in 1805, and was immediately offered a position teaching religion at the University of Prague where he had been a student. During the next fifteen years, he published several important mathematical results, in addition to his religious work.

Bolzano was a deeply principled man, and it was important to him that everything, including mathematics, rest on firm foundations. Although many great mathematicians had been using calculus and expanding its applications in the century since it had been invented, it was based on the poorly-understood concept of infinitesimals. Bolzano invented the ϵ-δ definition of continuous functions, and proved many fundamental results such as the intermediate value theorem for an arbitrary continuous function, commonly known as Bolzano-Cauchy. He also proved what is now called the Bolzano-Weierstrass Theorem, which says that an infinite sequence of points in an n-dimensional cube has a subsequence that converges to a limit — and he did this decades before Weierstrass did so.

During this period, Bolzano also taught and preached his ethical principles, which included the idea that social and economic inequalities were unjust. In an 1813 sermon, he said:

> There will be a time when the thousandfold distinctions of rank among men, which cause so much harm, will be reduced to the proper degree, when each will treat the other as a brother. There will be a time when constitutions will be introduced which are not subject to the same terrible abuse as the present one.

He also opposed war, militarism, and the glorification of violence. Bolzano's weekly sermons became very popular, and were regularly attended by hundreds of prominent citizens of Prague.

Needless to say, Bolzano's radical ideas did not sit well with the government. At first, the imperial authorities in Vienna were too busy fighting Napoleon's armies to pay too much attention. But after Napoleon's defeat in 1815, they started to take more notice. Finally, in 1819, Bolzano was fired from his position at the University of Prague and forbidden to teach or preach in public. He was also prevented from publishing his work in most major outlets. Bolzano never renounced his beliefs, but he did cleverly apologize for "the evil consequences that might have resulted from them being misunderstood."

Fortunately, Bolzano had patrons who continued to support him, encouraged his work, and helped get some of it published after his death. His book *Theory of Science*, in which he tried to provide logical foundations for all scientific knowledge, was published in 1837, and *Paradoxes of the Infinite* (discussed above), which he started writing at age 66, was published posthumously in 1851.

Bolzano died in 1848 in obscurity, his work unknown to the mathematical community. It was not until nearly 100 years later that his contributions were finally recognized. His collected works fill 39 volumes, most of which have never been translated from German. Because of the restrictions on his ability to publish, he may have made many more great discoveries that we'll never know.

In his book *Theory of Science*, Bolzano moved from sets to ideas about what we can and cannot prove, a transition we will also make in this journey. He writes:

> That no proposition has truth disproves itself because it is itself a proposition and we should have to call it false in order to call it true. For, if all propositions were false, then this proposition itself, namely that all propositions are false, would be false. Thus, not all propositions are false, but there also true propositions.

As we shall see, this idea of self-referential propositions will play an important role in the development of set theory and computability.

Exercise 3.1. Using Bolzano's method, prove that there are infinitely many true propositions.

4 Set Theory

In 1873, Georg Cantor, a professor at the University of Halle in Germany, had an important correspondence with Richard Dedekind, about the size of various sets. Cantor asked whether it was possible to enumerate positive real numbers. Dedekind replied that this was "not an interesting question," but provided a proof of how to enumerate algebraic numbers (we'll explain what these are in a minute).[4] Cantor then said that if the answer to his original question is no, it would provide a new and more powerful proof of the existence of transcendental numbers. Then, in his next letter, Cantor announced that he had proved that the answer was indeed no.

To understand Cantor's proof, it's important to understand the notion of *countable* and *uncountable* sets.

Definition 4.1. *Countable sets are sets that can be put into a one-to-one correspondence with natural numbers.*

As we have seen, Oresme and Bolzano were already using this idea to show that various sets were "the same size" as integers. Hilbert later came up with a nice way to think about these sets, which we now call Hilbert's Hotel: Imagine that we have a hotel with an infinite number of rooms, each holding one guest. Is the hotel full, or can it accommodate more guests?

- Suppose one guest arrives at the hotel. Is there room for her? Yes; the hotel simply asks the guest in room 1 to move to room 2, the guest in room 2 to move to room 3, and so on. The new guest can then move into the now-vacant room 1.
- Suppose infinitely many new guests arrive. Is there room for all of them? Yes; the hotel asks the guest in room 1 to move to room 2, the guest in room 2 to move to room 4, and so on. Now all the odd-numbered rooms are vacant. Since there are infinitely many odd numbers, the new guests will all have rooms.
- Suppose infinitely many buses, each holding infinitely many guests, arrive. Is there room for them? Yes! The odd-numbered rooms are emptied as above. Then the people in the first bus move into rooms numbered 3^i, the people in the second bus move into rooms numbered 5^i, and so on for each prime number.[5]

What about rational numbers? At first glance it might seem like there are more rational numbers than integers, since there are infinitely many of them between each integer. Yet as with the infinite number of infinite buses, it turns out that there is a way to count them all with integers. Imagine that we have an infinite grid of rational numbers, starting with

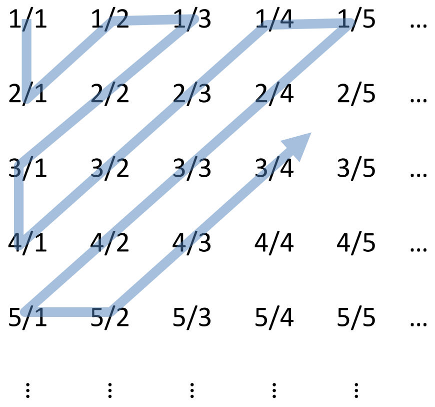
[4] While "everyday" numbers are algebraic, it turns out that they account for only a tiny amount of all numbers; the rest are transcendental numbers like π.

[5] There are actually many ways to solve this case.

$\frac{1}{1}$. As we move from one column to the next, we'll increase the denominator by 1; as we move from one row to the next, we'll increase the numerator by 1, so the upper left corner of the grid will look like this:

1/1	1/2	1/3	1/4	1/5	...
2/1	2/2	2/3	2/4	2/5	...
3/1	3/2	3/3	3/4	3/5	...
4/1	4/2	4/3	4/4	4/5	...
5/1	5/2	5/3	5/4	5/5	...
⋮	⋮	⋮	⋮	⋮	

To count all the rational numbers, start at the upper left corner and follow this zigzag pattern:



1/1	1/2	1/3	1/4	1/5	...
2/1	2/2	2/3	2/4	2/5	...
3/1	3/2	3/3	3/4	3/5	...
4/1	4/2	4/3	4/4	4/5	...
5/1	5/2	5/3	5/4	5/5	...
⋮	⋮	⋮	⋮	⋮	

Since we can put rational numbers in a one-to-one correspondence[6] with natural numbers, rational numbers are also a countably infinite set.

[6] Actually, we need to first drop reducible pairs from the table to make a one-to-one correspondence.

A real number is called *algebraic* if it is a root of a polynomial with integer coefficients. A polynomial is said to have *weight* k if k is the maximum of the absolute values of its coefficients and exponents. Since there is a finite number of polynomials of a given weight, and every one of them has finitely many roots, algebraic numbers are countable. Although people think of this as one of Cantor's proofs, it actually came from Dedekind — it's the one he mentioned in the letters described above. (Interestingly, it was Emmy Noether who published their correspondence.)

Cantor's 1874 Theorem: Uncountability of Continuum. Now we come to Cantor's theorem that the set of all real numbers (also called the *continuum*) is uncountable:

> Given a closed interval of real numbers and a sequence of real numbers, the interval contains a number that is not in the sequence.

Formally, we can state it like this:

$$\text{If } a_0, b_0 \in \mathbb{R} \ \land \ a_0 < b_0 \text{ and } \{r_i\}_{i=0}^{\infty} \subseteq \mathbb{R}$$
$$\text{then } \exists c \in [a_0, b_0] \text{ such that } c \notin \{r_i\}.$$

Informally, what it says is that you cannot enumerate every point in a line segment. The idea of the proof is to imagine that there is a list of all the real numbers, and then show that there is a number that is not on the list.

Proof:

1. Find the first pair of distinct elements in the sequence $\{r_i\}$ that is in the interval $[a_0, b_0]$. (If there is no such pair, then, obviously there is an element in $[a_0, b_0]$ that is not in the sequence; therefore; we are done with the proof.)

2. Designate the smaller element of the pair a_1 and the larger one b_1, and repeat steps 1 and 2 with the interval $[a_0, b_0]$ being replaced with the $[a_1, b_1]$ and the sequence $\{r_i\}$ with the "unused" elements in the sequence $\{r_i\}$.

3. Observe that r_i does not belong to the inside of the interval $[a_{2i}, b_{2i}]$.

4. So if the sequence of intervals $[a_i, b_i]$ is finite, the middle of the last interval is not in the sequence $\{r_i\}$.

5. If the sequence of intervals $[a_i, b_i]$ is infinite, no element r_i could belong to its intersection $\bigcap[a_i, b_i]$. And since we know that the intersection is not empty, an element from it will not be in $\{r_i\}$.

Exercise 4.1. Prove that the intersection of nested, closed intervals is not empty.

Georg Cantor (1845–1918)

Georg Cantor was born in Russia to German parents. His family moved back to Germany when he was a boy, and he studied first at the University of Zurich and then at the University of Berlin. He received a Ph.D. in 1867, working on number theory. He then hoped to be hired at one of the top research universities, but the only teaching job he found was at the less prestigious University of Halle.

During the 1870s and 1880s, Cantor developed most of his important work on set theory. He was greatly influenced by Bolzano's *Paradoxes of the Infinite*, and was also assisted by Richard Dedekind, with whom he kept up a long correspondence refining his ideas.

Cantor's work on infinite sets was so radical that other mathematicians at the time dismissed and ridiculed it. Poincaré thought all of set theory should be thrown out of mathematics. Leopold Kronecker, a prominent professor at the University of Berlin, insured that Cantor never got an appointment there; Kronecker reportedly said that he had to "protect the youth from the deviations of Cantor."

Yet Cantor believed that his infinite sets were not only a great mathematical discovery, but a bridge between humanity and God. He wrote letters to Pope Leo XIII and to a famous theologian, Cardinal Franzelin, telling them that his work on infinities had finally proven the existence of God. Devastated by the reception of his ideas, he withdrew into depression and periodically spent time in mental hospitals for the rest of his life. For a while, he abandoned mathematics and attempted to prove that Francis Bacon wrote Shakespeare's plays. Despite the gradual acceptance of his ideas in the mathematical community, Cantor was dejected for many years and died lonely and in poverty.

Despite his apparent failure, Cantor's work actually ended up revolutionizing mathematics. By about 1930, the idea that set theory provides a foundation for all of mathematics had become widely accepted, and remains so today. Cantor changed the criteria for existence of a mathematical entity: for him (and mathematicians today), to show that something exists, it is enough to show that it does not introduce any logical contradictions.

Cantor realized that there were two kinds of infinity. One was the cardinality ("size") of the natural numbers $\mathbb{N}$ and other sets equipotent to (i.e. that could be placed in a one-to-one correspondence with) $\mathbb{N}$. He labeled this number $\aleph_0$, using the first letter of the Hebrew alphabet, *aleph*. The second was the cardinality of sets equipotent to the real numbers $\mathbb{R}$, which he called $\mathfrak{C}$ (the letter C, for "continuum," in a Gothic typeface).

Another of Cantor's important results concerned the size of the *power set* of the set of natural numbers. The power set of a set S, written 2^S or $\mathcal{P}(S)$, is the set of all subsets of S:

$$x \in 2^S \iff x \subseteq S$$

The notation derives from the fact that if S is finite, it has $2^{|S|}$ possible subsets.

Exercise 4.2. Prove that a power set of a finite set with n elements contains 2^n elements.

Cantor saw that the power set of $\aleph_0$ had the same cardinality as the real numbers:

$$|2^{\aleph_0}| = |\mathfrak{C}|$$

This followed from the fact that one set could be mapped to the other, following several steps:

1. $|\mathbb{R}| = |\mathbb{R}^+|$ through $f(x) = e^x$. That is, we can map all real numbers to just positive reals by raising e to the power of the number.
2. $|\mathbb{R}^+| = |(0,1)|$ through $f(x) = \frac{1}{x+1}$. We can map positive reals to the open interval between 0 and 1 by applying the given function.
3. $|(0,1)| = |\text{Seq}\{0,1\}|$ through binary. We can map any point in this interval to a infinite sequence of binary digits, representing a binary fraction (imagine a "binary point" before the first digit).[7] We denote the set of all such sequences Seq{0, 1}.
4. $|\text{Seq}\{0,1\}| = |2^{\aleph_0}|$ through the *characteristic function*. This is similar to the way we use bit vectors in programming. That is, we can use the first binary digit to indicate whether the first possible element is present in a given subset, the second digit to indicate whether the second possible element is present, and so on. [8]

5 Diagonalization

In 1903, Bertrand Russell stated what is now called *Russell's Paradox*: Consider the set $S = \{x \mid x \notin x\}$ of all sets that do not contain themselves. Is $S \in S$? If it is, it is not; if it is not, it is. This idea was similar in spirit to a proof technique Cantor developed several years earlier, when proving perhaps his most important result. Published late in his career in 1891, it is now known as *Cantor's Theorem*:

[7] In fact, the situation is slightly more complicated, because there is more than one way to represent the same real. For example, $\frac{1}{2}$ could be represented in binary as 0.1, or as the infinitely repeating 0.01111.... However, there are ways to address this ambiguity.

[8] For example, the set of odd numbers would be represented by the bit string 010101... because the first natural number (0) is not in the set, the next one is, the next one is not, and so on.

There is no onto function from a set to its power set.

An onto function is one that covers all the values in its codomain. Formally, a function $f : X \to Y$ is called *onto* or *surjective* if

$$\forall y \in Y \; \exists x \in X \; : \; f(x) = y$$

Note that a function can be onto without being one-to-one (also known as *bijective*). For example, the absolute value function from integers to natural numbers is onto (every natural number is the absolute value of some integer) but not one-to-one (since two different things, an integer and its negative, both have the same absolute value).

In proving his theorem, Cantor introduced a method known as a *diagonal argument* or *diagonalization*, which has since become one of the standard proof techniques used by mathematicians.

Here's Cantor's proof:

> Assume the contrary, i.e. that there exists a set $\mathbb{S}$ and a surjection $f : \mathbb{S} \to 2^{\mathbb{S}}$.
>
> Define a set $\mathbb{D}$ consisting only of elements of $\mathbb{S}$ that are not in the image of the mapping:
> $$\mathbb{D} = \{x \in \mathbb{S} \mid x \notin f(x)\}$$
> Since $\mathbb{D} \subset \mathbb{S}$, then by definition of power set, $\mathbb{D} \in 2^{\mathbb{S}}$.
> Since f is surjective, every element of $2^{\mathbb{S}}$, including $\mathbb{D}$, must have at least one element in $\mathbb{S}$ that maps to it. Let's call d the element that maps to $\mathbb{D}$. In other words, $f(d) = \mathbb{D}$.
> Either $d \in \mathbb{D}$ or $d \notin \mathbb{D}$.
> Suppose $d \in \mathbb{D}$. Then by the definition of $\mathbb{D}$, $d \notin f(d)$. But we chose d so that $f(d) = \mathbb{D}$, so this means $d \notin \mathbb{D}$, a contradiction.
> On the other hand, suppose $d \notin \mathbb{D}$. Since $f(d) = \mathbb{D}$, this means $d \notin f(d)$. But then by the definition of $\mathbb{D}$, $d \in \mathbb{D}$, also a contradiction.
> So our initial assumption must be wrong, and there can be no onto function from a set to its power set.

To see why this approach is called diagonalization, imagine that there's a way to write down all the subsets of a set S. Of course, S might be (countably) infinite, so our list might be infinite, but that's okay. The way we'll write each subset is with its characteristic function: a binary string with a 1 in position i if the ith element is in this subset, and a 0 otherwise. So our list might look like this:

		ELEMENT #					
		1	2	3	4	5	...
SUBSET #	1	1	1	1	1	1	...
	2	1	0	1	0	1	...
	3	0	1	1	0	1	
	4	0	0	1	1	0	...
	5	1	1	0	0	1	...
		⋮	⋮	⋮	⋮	⋮	

To construct our subset D, we want to identify a new subset d that's not already on the list. How do we do this? We take the bit representing the presence or absence of the first element of the first subset and negate it — if it's a 1, then d will have a 0 in that position; if it's 0, then d will have a 1. Then we do the same for the second element of the second subset, and so on, moving along the diagonal:

		1	2	3	4	5	...
SUBSET #	1	**1**	1	1	1	1	...
	2	1	**0**	1	0	1	...
	3	0	1	**1**	0	1	
	4	0	0	1	**1**	0	...
	5	1	1	0	0	**1**	...
		⋮	⋮	⋮	⋮	⋮	

How do we know that d isn't already on the list? Well, we know it's not the same as the first entry on the list, because that entry contains the first element of S (i.e., has a 1 in the first position), while d does not. We know it's not the same as the second entry on the list, because that entry has a 0 in the second position, while d has a a 1. Continuing down, we see that d can't be on the list because it *always* differs from every list entry in at least one position — the one on the diagonal.

In fact, we can use this diagonal argument to prove the uncountability of the continuum much more simply than Cantor did in his earlier 1874 proof. We just treat the bits as the binary representation of a number in the interval (0,1). However many entries we add to the table, we can always use the above diagonal construction to find a number d that is not on the list.[9]

6 The Continuum Hypothesis

We know that $\aleph_0$ is the smallest kind of infinity, and that $\mathfrak{C}$ is much bigger. But are there other levels of infinity in between? Is there an $\aleph_1$ that is bigger than $\aleph_0$ and smaller than $\mathfrak{C}$? Cantor believed that there was not; in other words:

$$\mathfrak{C} = 2^{\aleph_0} = \aleph_1$$

This claim, which Cantor stated in 1878, is called the *continuum hypothesis* (*CH* for short), and he spent a great deal of effort trying and failing to prove it.

Hilbert made the problem famous in 1900 when he listed it as the first of his 23 problems. Mathematicians worked on it for decades, ultimately showing that it cannot be disproved (1940 by Kurt Gödel) or proved (1963 by Paul Cohen) within standard set theory.[10]

Whether or not there are infinities in between $\aleph_0$ and $2^{\aleph_0}$, Cantor's Theorem tells us that there are bigger ones, each formed by taking the power set of elements of the previous one: $\aleph_0, 2^{\aleph_0}, 2^{2^{\aleph_0}}, \ldots$. We call the ith element of this sequence $\beth_i$ (using the Hebrew letter *beth*). Of course, this sequence is itself infinite, and we could take the union of these sets as i grows without bound:

$$\bigcup_{i=0}^{\infty} \beth_i$$

These are very strange (and very large) "unreachable" infinities. In a 1952 book called *Les Nombres Inaccessibles* ("Inaccessible Numbers"), Émile Borel pointed out that since sentences of a natural language are countably infinite, then only a tiny subset of all numbers can be referred to with natural language sentences. So most numbers are "inaccessible"

[9] In fact, we don't even need to represent the interval in binary. We could represent the numbers on the list as decimal numbers with an implicit decimal point before them; instead of negating the ith digit, we simply add 1 mod 10, or use some other function for getting a different value. All that matters is that we make sure the ith digit of the number we're constructing is different from the one used in the ith number in the table.

[10] "Standard set theory" here refers to what are known as ZF and ZFC, which will be explained in the next section.

in the sense that they can never be referred to — in fact, if you take a random real number in the interval [0,1], it will be inaccessible with probability 1! This raises an interesting philosophical question: In what sense do these numbers exist?

7 Axiomatizing Set Theory

In 1907, mathematician Ernst Zermelo proposed seven axioms to create a solid foundation for set theory, with no paradoxes. We'll look at each axiom and comment on it:

> I. *Axiom of extensionality.* If every element of set M is also an element of N and the other way around, then $M = N$.

It's important to define what it means for sets to be equal. It's hard, because sets are unordered collections. In programming, it's very hard to implement sets unless you assume some total ordering.

> II. *Axiom of elementary sets.* There is a set, the *empty set* $\varnothing$, that contains no elements. If a is an object of the domain, there exists a set $\{a\}$, that contains a and only a as an element. If a and b are two objects of the domain, there always exists a set $\{a, b\}$ containing as elements a and b but no object x distinct from them both.

This is analogous to the way the Lisp programming language builds up everything from `nil` and the `cons` operation.

> III. *Axiom of separation.* Whenever the propositional function $E(x)$ is defined for all elements of a set M, M possesses a subset containing all the elements x of M for which $E(x)$ is true and no other elements.

This allows you to get a subset out of a set.

> IV. *Axiom of the power set.* To every set T there corresponds a set $\mathcal{P}(T)$, the power set of T, that contains all the subsets of T and no other elements.

For Zermelo, it's only a set if it can be built up from $\varnothing$ by repeated applications of power set: $2^{\varnothing} = \{\varnothing\}, 2^{2^{\varnothing}} = \{\varnothing, \{\varnothing\}\}$, and so on.

> V. *Axiom of the union.* To every set T there corresponds a set $\cup_T$, the union of T, that contains all the elements of the elements of T and no other elements.

This "flattens" one level of a nested set.

> VI. *Axiom of Choice (AC).* If T is a set whose elements all are sets that are different from $\varnothing$, and mutually disjoint, its union $\cup_T$ includes at least one subset C_T having one and only one element in common with each element of T.

This is a very peculiar axiom; it says you can pick an element. If you think of all the subsets as bags, C_T is like getting one element from every bag. It turns out that this has some very strange consequences, such as the *Banach-Tarski* paradox, which says that by using the axiom of choice, one can cut a sphere into a finite number of pieces that can be rearranged so that we end up with two spheres of the same size as the original sphere. Because of these problems, many mathematicians have wondered if it's possible to get rid of this axiom.

> VII. *Axiom of infinity.* There exists in the domain at least one set Z that contains the empty set as an element and is so constituted that to each of its elements a there corresponds a further element of the form $\{a\}$.

When talking about infinite sets, most people today use Dedekind's axiom that it is a set having a one-to-one correspondence with a subset. But Zermelo used a constructive method, defining infinite sets recursively. The use of recursion in the definition of data structures (again, common in Lisp) dates back to his idea.

Exercise 7.1. Prove the following theorem due to Zermelo: Every set M possesses at least one subset M_0 that is not an element of M.

In 1925, Israeli mathematician Abraham Fraenkel proposed two additional axioms:

> VIII. *Axiom of regularity.* Every non-empty set contains an element disjoint from it.
>
> IX. *Axiom of replacement.* An image of every set is a set.

These axioms help mathematicians deal with infinities of infinities, and we won't be discussing them further. What's important is that the resulting combined theory, known as *Zermelo-Fraenkel set theory with the axiom of Choice (ZFC)* became the generally accepted approach to providing a foundation for mathematics. (Without the axiom of choice, the theory is known as ZF.)

Like non-Euclidean geometry, set theory progressed from being considered a crazy idea to being accepted as a standard part of mathematics. When Cantor invented set theory in the 1870s, he was ridiculed and considered to be a threat to youth. Twenty years later, the top researchers in the world were working on set theory, and by the 1920s, it was generally considered to provide a foundation for mathematics. What's interesting is that nothing changed about the theory. It's not that the problems were addressed or that the paradoxes became any less paradoxical. What changed was the tenor of the times; ideas that seemed radical in one century were perfectly acceptable in another. We see this pattern throughout the history of mathematics — and through history in general.

Now we'll see how attempts to formalize mathematics led to some striking results about the abilities and limits of any computational device.

8 Hilbert's Program

In 1921, David Hilbert proposed that all mathematics be formalized, which meant:

- Every proposition is written in a formal language.
- The system must be *complete*, that is, every true proposition can be proved within the system. If a proposition is not true, then its negation is provable.
- The system must be *consistent*, meaning that it is not possible to prove both a proposition and its negation.
- Completeness and consistency should be proven with "finitary" methods, that is, without resorting to infinities.

This became known as *Hilbert's Program*, and it became a focus of work for many important mathematicians in the 1920s.

What Hilbert was essentially proposing was a system for mechanically proving or disproving theorems. This system could be thought of as a (possibly abstract) computational machine, in which one could simply "turn the crank" and the system would decide whether a proposition was true or false.

As time went on, there was substantial progress. For example, in 1929, Kurt Gödel proved that First-Order Predicate Calculus (FOPC),[11] a formal system that could be used to prove some mathematical propositions, was complete. By 1930, almost everyone in the mathematical world believed that Hilbert's Program would be accomplished at any moment.

9 The Program Collapses

> "It is all over."
>
> — John von Neumann, on hearing about Gödel's Theorem

[11] FOPC contains predicates that take individual elements as arguments; for example, we might have a predicate $P(x)$ that is true when x is a person but false otherwise. "First-Order" means that we can have predicates about elements, but we can't have predicates about predicates.

Gödel began to focus his work on the system proposed by Alfred North Whitehead and Bertrand Russell in their three-volume work, *Principia Mathematica*[12] (PM). The book, whose final volume came out in 1913, was an earlier attempt to formalize mathematics. The system described in PM was more powerful than FOPC — powerful enough to satisfy Hilbert's first requirement. But was it complete?

Gödel realized that PM's use of induction, which FOPC lacks, makes all the difference. It makes the system more powerful, but it also introduces a weakness. In a 1931 paper "On Formally Undecidable Propositions in *Principia Mathematica* and Related Systems I," Gödel published his famous *incompleteness theorems*, which stated that:

1. Any consistent theory containing Peano arithmetic contains a proposition that is true, but not provable.
2. Any consistent theory containing Peano arithmetic cannot prove its own consistency.

Essentially, what these two results said is that Hilbert's program was fundamentally impossible. Any system strong enough to express the kind of mathematical propositions needed to do arithmetic was necessarily *incomplete*.

Gödel's proof is quite complex, but the basic idea is this: The axioms and inference rules in PM are designed to prove propositions about integers. For example, it can express, in formal symbols, an idea like "5 is prime" or "there are infinitely many even numbers." So Gödel came up with a clever way of assigning a unique integer to every possible proposition, a technique now known as *Gödel-numbering* or *Gödelization*. Since propositions in the system are about numbers, and propositions could now themselves be expressed as numbers, it was possible to have propositions that "talk about" the truth and falsity of other propositions. Gödel then showed how to express a special proposition in the language of the system, a proposition which essentially says:

> *This proposition is not provable.*

If this proposition could be proved mechanically within the system, then the statement is false, so the system has proved a false proposition, and is therefore inconsistent. If the proposition can't be proved by the system, then it is a true statement that can't be proven, and the system is incomplete.

It is difficult to convey today how shocking Gödel's result was. Prior to 1931, there was a general sense that all mathematical truths, and perhaps eventually all scientific truths, would ultimately be derivable by a mechanistic process, and that researchers were on the verge of doing so. After Gödel's Theorem, these hopes were dashed. As the quote

[12] Although Whitehead and Russell were English, they gave their book a Latin title, meaning "Mathematical Principles," probably named after Isaac Newton's similarly named book from 200 years earlier. Confusingly, Russell also wrote an entirely different book called *Principles of Mathematics*.

from von Neumann at the beginning of this section suggests, there was a feeling that the foundations of their intellectual enterprise were crumbling.

Yet decades later, we are more comfortable with the idea of these limits. Mathematics existed before anyone tried to formalize its foundations, and continued to exist afterwards. No bridges collapsed because of Gödel's result.

Interestingly, the idea of Gödel-numbering, which was so novel in 1931, is quite familiar to computer scientists today: We know that a string of bits can represent a number or an instruction in machine language, and the instruction can act on other numbers or even on itself.

Kurt Gödel (1906–1978)

Kurt Gödel (pronounced almost like the English word "girdle," but without finishing the "r" sound) was born in Brno, in what is now the Czech Republic but was then the Austro-Hungarian empire. He came from an upper-middle class ethnically German family; he always considered himself Austrian.

In 1924, Gödel moved to Vienna to attend the university there. Vienna in the 1920s and 1930s was a hotbed of radical new thinking. Painters were moving from representational to abstract art; composers like Schoenberg were rejecting tonality and other conventions from traditional music. Artists in general were focused on challenging and rethinking the foundations of their disciplines and no longer cared whether they pleased an audience.

In Philosophy, a group called the Vienna Circle formed in response to the early work of fellow Viennese Ludwig Wittgenstein. They looked at how to analyze the logical structure of language of science, and manipulate these logical propositions as a kind of game, rather than as an explanation of reality. Gödel was heavily influenced by this group, and was allowed to participate in their meetings though he was only a graduate student.

In 1928, Hilbert and Ackermann developed First-Order Predicate Calculus (FOPC) as part of their work on the foundations of mathematics, and Gödel became interested in this logical system. In 1929, for his doctoral thesis, Gödel was able to prove that FOPC was complete. It is possible that he hoped to extend this result to Russell and Whitehead's metamathematical system. In any case, as we saw above, he ultimately proved the reverse — that no such system could be complete — in his famous 1931 incompleteness theorems.

For the next few years, Gödel frequently traveled to Princeton University, lecturing about his work, which led to the creation of the new field of mathematical logic. He became friends with logicians Alonzo Church and Stephen Kleene as well as Albert Einstein.

In 1938, after Nazi Germany annexed Austria, Gödel decided to move to the United States. He joined the Institute for Advanced Study at Princeton where he had previously visited, and eventually became a U.S. citizen.

At first, he continued to do productive work. In 1940 he proved one of the most important results in set theory, showing that the continuum hypothesis (see Sec. 6) and the axiom of choice (see Sec. 7) are consistent with Zermelo-Fraenkel axioms. In other words, he showed that adding these to the initial axioms would not lead to a contradiction.

However, after this his mathematical work dried up. He never published another paper after the 1940s, despite living until age 71, and wrestled with borderline psychological problems. He spent much of his later years trying to logically prove the existence of God. He also became obsessed with the idea that his food might be poisoned, and refused to eat anything not prepared by his wife. When she was hospitalized for a serious illness for several months, Gödel gradually starved to death in 1978.

It is ironic that Gödel's tragic death was a result of the very actions he took trying to avoid the possibility of dying, a paradox reminiscent of the logic of his own incompleteness theorem.

10 Church, Turing, and the Origins of Computer Science

Is there a mechanical procedure for deciding whether a mathematical statement is true or false? This question, called the "decision problem" (*entscheidungsproblem* in German), is closely related to Hilbert's program, and it was the focus of many researchers in the 1930s.

One of these was Alonzo Church. Church came up with a system for defining computational procedures, which he called *λ-calculus*. And he formulated an important hypothesis, known as *Church's Thesis*, about the nature of computation:

> *Whatever can be computed by a modern computer (or by a Turing machine) includes everything that ever can be computed.*

Church didn't actually state it in this way; early versions of the idea referred to the kinds of abstract computational mechanisms that had been proposed at the time, like λ-calculus. But the basic idea is the same: once you have a basic mechanism for computation, you can (theoretically) compute everything computable. Every computational device is in some sense the same; there isn't some special kind of über-computer that can compute "harder" functions.[13]

Of course, Church's Thesis isn't provable, because we'd have to quantify over all possible future computational systems that don't exist yet. Nevertheless, it is generally accepted, and is in a sense the basis of Computer Science: We know that if we design an algorithm, we can implement it on any general-purpose computer.

[13] We're assuming that we have as much memory as we need and we don't care how long it takes.

Alonzo Church (1903–1995)

While many of the great mathematicians we've met were tragic figures, Alonzo Church was not one of them. With the exception of a childhood BB gun accident that left him blind in one eye, Church had a happy life. He enjoyed spending time with his wife and children, and had a wonderful career spanning much of the 20th century.

Church grew up in Virginia, then attended college at Princeton University, where several members of his family had gone. He published his first paper while still an undergraduate. After graduating with a degree in Mathematics, he stayed for a Ph.D., which he got in 1927 after just three years.

Church spent the next two years doing research at several top universities. He visited both the University of Göttingen (home of David Hilbert and the formalist approach) and the University of Amsterdam (home of L.E.J. Brouwer, founder of the intuitionist approach).

in 1929, Church returned to Princeton to join the faculty. Over the next few years, he began gathering a group of students working on what he realized was a new area of mathematics, which is called *mathematical logic* or *symbolic logic.* Kurt Gödel came to lecture; Alan Turing heard about Church's work and came to Princeton to finish his Ph.D. with Church as his advisor.

During this period, Church did more than anyone else to establish and promote this new field. First, he published *A Bibliography of Symbolic Logic*, identifying the most important writings. By doing so, Church defined what the new field was about. He also co-founded the Association for Symbolic Logic, and served as its first president. Next, he founded the *Journal of Symbolic Logic*, which still exists today. He decided to become editor of the reviews section, insuring for the next 40 years that a review of every important book in the field appear in the journal.

Church taught courses in symbolic logic, and in 1944, published the first volume of *Introduction to Mathematical Logic*, which became the definitive textbook in the field for decades. Unfortunately, he never completed a planned second volume, which would have included a chapter on Gödel's Theorem and other advanced topics.

Perhaps Church's greatest legacy is in his Ph.D. students, a group which includes not only leaders of the field of symbolic logic, but many who went on to be important pioneers in Computer Science, including two Turing Award winners — not to mention Turing himself.

Church moved to UCLA in 1967, after they promised to continue supporting the editorial staff of the journal. Although he was already 64 at the time, he continued to teach until age 87, and was still publishing papers until his death in 1995 at age 92.

Independently of Church, British mathematician Alan Turing developed his own approach to the *entscheidungsproblem*. In a 1937 paper, Turing proposed the idea of an abstract computational machine — what we now call a *Turing machine* — that reads data from a tape and takes actions based on a set of rules. He also recognized that a *universal* version of this machine could, with suitable input, simulate the function of any other machine. Turing proved that given an arbitrary program, there was no way to decide whether the machine would halt when running the program. Since this *halting problem* was undecidable, the could be no solution to Hilbert's more general decision problem.

The Turing machine provided a critical bridge between mathematics and computer science. While Gödel and Church had created their own mathematical formalisms for describing computable functions, Turing was the first to do so with a (hypothetical) mechanical device. Indeed, although the Turing machine was never meant to be a physical device, its ideas greatly influenced von Neumann and others in their designs for actual general-purpose computers.

Turing later studied with Church, and their early work overlapped a great deal. Today we refer to the devices in Church's Thesis as "Turing-complete" — that is, computational systems equivalent in power to a Turing machine. (In fact, Church's Thesis is often called the *Church-Turing Thesis* due to Turing's contributions.)

Alan Turing (1912–1954)

Alan Turing grew up in England, had a respectable boarding-school education, and attended King's College at Cambridge, where he graduated with a degree in mathematics in 1934. He continued his studies there as a fellow, and it was during this two-year period that he developed the notion of a Turing machine. Turing's paper describing these results was submitted in 1936, and published a few months later. He then went to Princeton University to study with Alonzo Church; his 1938 Ph.D. thesis at Princeton involved further extensions to ideas of computability.

In 1938, with World War II clearly approaching, Turing joined the British government's cryptographic agency, eventually moving to its headquarters in Bletchley Park for the duration of the war. Turing was one of the top minds at Bletchley, playing key roles in the development of mechanical devices for breaking Nazi codes.

After the war, Turing wrote a specification for a computer called the ACE (Automatic Computing Engine), which he hoped would be built by the government research lab where he worked. It was the most detailed proposal yet for a programmable digital computer. Some of the ideas for the ACE undoubtedly came from his experience designing digital devices at Bletchley, but because of the Official Secrets Act, he was forbidden from discussing any of that work or explaining how he knew his designs were feasible. As a result, there was relatively little support for funding construction of the ACE. While a preliminary version, the "Pilot ACE" was eventually built, the full ACE design was never implemented. Turing left in frustration and took a job at the University of Manchester in 1949.

During his time at Manchester, Turing worked on a variety of important topics in several fields. He developed the LU-decomposition, a commonly used technique in linear algebra. In later years, he worked on morphogenesis, trying to understand the mathematics behind the growth of plants. But perhaps his best known work from this period was the paper “Computing Machinery and Intelligence,” published in 1950 in the important philosophical journal *Mind*. In this paper, Turing addresses the question of whether a machine could ever think. He proposes that rather than arguing over the meaning of intelligence, we should instead construct a test (now known as the Turing test) in which a human interrogator tries to distinguish between a human and a computer by typing questions to each and reading the answers. He goes on to talk about possible objections to the test, and refutes them in various ways. Turing’s paper is still considered one of the foundational works in the field of Artificial Intelligence.

Despite these achievements, Turing’s time at Manchester was marred by an incident that dramatically affected his life, and possibly influenced his death. In 1952,Turing, who was gay, had his house burgled by a man with whom he’d had a casual sexual relationship. When he reported the crime to the police, they responded by arresting Turing himself for the crime of “gross indecency,” since homosexuality was still illegal in Britain. Although this was rarely prosecuted at the time, the local officials felt obligated to pursue the case against Turing. Despite his great contributions to his own country during the war, he was convicted. He was sentenced to an experimental “treatment”: a year’s worth of injections of synthetic estrogen, which caused a variety of unpleasant intended and unintended effects on his body and possibly his mental state.

In 1954, at age 41, Turing was found dead from cyanide poisoning, a half-eaten apple by his bed. For years it was assumed that he had committed suicide by dipping the apple in poison, but the apple was never tested. More recently, there has been some skepticism about that conclusion; it’s possible that his death was due to accidental inhalation of cyanide fumes.

Turing’s contributions to computing are unmatched, and today the highest award given in Computer Science is known as the Turing Award in his honor. We can only speculate about what else he might have done had he lived long enough to see computers become a part of everyday life.

11 Diagonalization Revisited

As discussed in Section 5, Cantor’s diagonal argument became the basis of a standard proof technique in mathematical logic and computer science. The argument relies on selfreference (specifically, in the case of Cantor's Theorem, in the definition of the set D as $\{x \in S | x \notin f(x)\}$).

Self-referential propositions have a long history, dating back at least 600 BC when the Cretan philosopher Epimenides said "All Cretans are liars."[14] Medieval philosophers wrestled with these so-called *insolubilia* ("unsolvables").

In Journey 2, we saw how we could start with a specific algorithm — greatest common measure of two line segments — and, through repeated abstraction, apply it to more and more general domains. Could we do the same thing for diagonalization? What would a generic form of Cantor's proof look like?

To do this, we'll need to start by defining a particular set of functions:

Definition 11.1. *A set $\mathbb{F}$ of (possibly partial) functions of one ($\mathbb{F}_1$) and two ($\mathbb{F}_2$) variables from domain T to codomain T' is called a* unary-binary family.

(Here, "unary" and "binary" refer to the *arity* of the function, i.e. how many arguments it has, not to unary or binary numbers.) Some examples of unary-binary families are:

- total functions from $\mathbb{N}$ to $\mathbb{N}$ and from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$
- continuous functions from $\mathbb{R}$ to $\mathbb{R}$ and from $\mathbb{R} \times \mathbb{R}$ to $\mathbb{R}$
- computable functions from $\mathbb{N}$ to $\mathbb{N}$ and from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$
- functions from $\mathbb{Z}$ to $\mathbb{Z}_2$ and from $\mathbb{Z} \times \mathbb{Z}$ to $\mathbb{Z}_2$

(Functions with more than two arguments may also be included in the family, but not needed for any of our results.)

Definition 11.2. *In a unary-binary family $\mathbb{F}$ with the domain T a binary function $\mathfrak{I} \in \mathbb{F}_2$ is called an* interpreter *if*

$$\forall f \in \mathbb{F}_1 \ \exists c \in T \ : \ \forall x \in T \ \mathfrak{I}(c, x) = f(x)$$

We can think of c as representing code, x representing the input data we're going to run the code on, and $\mathfrak{I}$ (the letter I in gothic typeface) as the interpreter. We normally think of an interpreter as something that takes some code and some data and executes the code on the data. However, here we are generalizing the notion of code (and the notion of interpreter); the code here is simply a way to indicate to the interpreter which function to compute. The interpreter can use any method it chooses to compute the function — executing instructions, using a lookup table, or even asking a genie. The code and data have to be the same type, but that's not a very stringent restriction — they could be bits, s-expressions in a Lisp-like language, and so on.

[14] Actually, Epimenides wasn't trying to state a paradox; in the original context, he was criticizing public opinion, and may not have realized the irony of his statement. In any case, his statement eventually came to be a canonical example of a liar paradox.

The programming language Lisp includes an `eval` function that evaluates an expression, applying the first element of the list to the arguments constituting the rest of the list. We can define our own two-argument version[15] of this that meets our criteria for an interpreter:

```
(defun eval1 (x y)
       (eval (cons x y)))
```

But our definition of interpreters doesn't require anything as powerful as this. Here's an example of a very simple interpreter: Consider a unary-binary family L containing unary functions $f_i(x) = x + i$ and binary functions $w_j(x, y) = x + y + j$. Then, $w_0(x, y) = x + y$ is an interpreter in L, since $w_0(i, x) = f_i(x)$.

Now we introduce a strange idea: an *anti*-interpreter. If an interpreter computes whatever its given code is supposed to do on its given data, an anti-interpreter computes *anything but* that. Let's define it formally:

Definition 11.3. *In a unary-binary family $\mathbb{F}$ with domain T, a function $\mathfrak{A} \in \mathbb{F}_2$ is called an* anti-interpreter *if*

$$\forall f \in \mathbb{F}_1 \;\; \exists c \in T \;:\; \forall x \in T \;\; \mathfrak{A}(c, x) \neq f(x)$$

($\mathfrak{A}$ is the gothic letter A, for anti-interpreter.)

Let's construct an anti-interpreter: As before, L is a unary-binary family containing unary functions $f_i(x) = x + i$ and binary functions $w_j(x, y) = x + y + j$. Then, $w_1(x, y) = x + y + 1$ is an anti-interpreter in L, since $w_1(i, x) = x + i + 1 = f_i(x) + 1 \neq f_i(x)$.

What we're saying here is that if our code was supposed to add i to its argument, we'll add $i + 1$, and therefore by definition we will be computing something other than what the code was supposed to compute.

Could we make an analogous anti-interpreter for Lisp based on `eval`? To answer the question, we need one more definition:

Definition 11.4. *A unary-binary family $\mathbb{D}$ from T to T' is called a* diagonalizable family *if*

$$\forall g \in \mathbb{D}_2 \;\; \exists f \in \mathbb{D}_1 \;:\; \forall x \in T \;\; g(x, x) = f(x)$$

In other words, a family of functions is diagonalizable if you can give the same value to both arguments and still end up with a function in the family. For example, Lisp functions are diagonalizable, because for any binary function `foo` we can define a function `bar`:

```
(defun bar (x) (foo x x))
```

This brings us to the *Anti-Interpreter Theorem*:

[15] Lisp has a similar two-argument function called `apply`, but its first argument is a function, not an s-expression.

A diagonalizable family does not have an anti-interpeter.

Proof: Let us assume that we have a diagonalizable family that *does* have an anti-interpreter $\mathfrak{A}$. Since our family is diagonalizable, we know that there's a function

$$\mathfrak{a}(x) = \mathfrak{A}(x, x)$$

in the family. Since $\mathfrak{A}$ is an anti-interpreter, by definition it gives "the wrong answer" for every function, and in particular, for our particular function $\mathfrak{a}(x)$:

$$\exists c \in T \ : \ \forall x \ \ \mathfrak{a}(x) \neq \mathfrak{A}(c, x)$$

But if we use the code c as our argument x, we have:

$$\mathfrak{a}(c) = \mathfrak{A}(c, c) \ \wedge \ \mathfrak{a}(c) \neq \mathfrak{A}(c, c)$$

Contradiction. So our assumption was impossible. Either the family is not diagonalizable, or it does not have an anti-interpreter.

What about our simple family L from p. 29? We've already shown that L has both an interpreter and an anti-interpreter. It is not diagonalizable because a function $w_k(x, x) = x + x + k = 2x + k$ is not in the family.

To come up with a general rule about when we have an interpreter, we need one more definition, a *composable-diagonalizable (C-D)* family:

Definition 11.5. *A diagonalizable family C with domain T and co-domain T' is* composable *if:*

$$\exists(\mathfrak{f} : T' \rightarrow T') \ \forall(w \in C) \ : \ \mathfrak{f}(x) \neq x \ \ \wedge \ \ \mathfrak{f}(w(x, y)) \in C$$

$\mathfrak{f}$ is called the *non-fixed-point function* of the family since it always returns something different from its argument. The definition says that the family is C-D if there is at least one non-fixed point function in the family that, when composed with any binary function in the family, is still in the family. It's a generalization of the "adding 1" idea we used above to make an anti-interpreter in our example family L.

Examples of C-D families include:

- Total computable functions over integers
- Polynomial time functions over integers
- Continuous functions over real numbers
- Differentiable functions over real numbers

In other words, they are exactly the kinds of functions we deal with every day.

Non-Existence of Interpreter Theorem: A C-D family does not contain an interpreter.

Proof: Let us assume that an interpreter $\mathfrak{I}$ exists. Then we can compose it with our function $\mathfrak{f}$. But the result will always give a different answer than $\mathfrak{I}$, so

$$\mathfrak{f}(\mathfrak{I}(x, y)) = \mathfrak{A}(x, y)$$

is an anti-interpreter. But we have already proven that a diagonalizable family can't have an anti-interpreter, because this will lead to contradiction.

Recall that the universal Turing machine is a Turing machine that is an interpreter for Turing machines. If we accept Church's thesis, such an interpreter exists in any equivalent formulation:

Interpreter Thesis: Any non-trivial set of computable functions that contains an interpreter is Turing-complete.

Using the interpreter idea, we can easily prove Turing's halting problem:

> *There is no computable function* halt(c, x) *that returns true if code c terminates on input x and false otherwise.*

Proof: If such a function existed, we would be able use it to construct an anti-interpreter:

```
integer anti_interpreter(integer c,  integer x) {
  if (halt(c, x)) {
    return interpreter(c, x) + 1;
  } else {
    return 0;
  }
}
```

But we have proven that an anti-interpreter cannot exist. So the function `halt` cannot exist.

Further Reading

Readers who are interested in learning more about these topics may wish to look at some of the references mentioned below. Complete citations are included in the Bibliography.

Infinity Before Cantor. A good overview of Oresme's work is Marshall Clagett's article "Nicole Oresme" in the *Dictionary of Scientific Biography*. (Despite the name, this is actually a large multi-volume encyclopedia.) Galileo's *Two New Sciences* contains only a few observations about infinities, but overall is a fascinating and beautifully-written book. Readers may be able to find this in the Galileo volume of Brittanica's *Great Books of the Western World* for under $10. Bolzano's work on set theory and logic is discussed in his book *Theory of Science*, translated and abridged by Rolf George.

Set Theory. A standard undergraduate textbook is Paul Halmos' *Naive Set Theory*. Cantor's original work on set theory is in *Contributions to the Founding of the Theory of Transfinite Numbers*. The comprehensive treatment of the topic of axioms of set theory is *Foundations of Set Theory* by Fraenkel, Bar-Hillel, and Levy. This is a remarkably accessible book considering its rigor and depth.

Logic and Computability. The best way to learn about Hilbert's program, Gödel's Incompleteness Theorems, and the foundations of computability theory is by reading the original papers. These are found in two books. The earlier works are in *From Frege to Gödel* edited by Jean van Heijenoort; the later ones (including those by Gödel, Church, and Turing) are in *Computability and Unsolvability*, edited by Martin Davis.

Alan Turing. An excellent biography is *Alan Turing: The Enigma* by Andrew Hodges.

References

[1] Bernard Bolzano and Rolf George (trans.). *Theory of Science: Attempt at a Detailed and in the Main Novel Exposition of Logic, with Constant Attention to Earlier Authors*. Blackwell, 1972.

[2] Georg Cantor. *Contributions to the Founding of the Theory of Transfinite Numbers*. Dover Publications, 1955.

[3] Marshall Clagett. Nicole Oresme. In Charles C. Gillespie, editor, *Dictionary of Scientific Biography*, volume 10, pages 223–230. Scribners, 1975.

[4] Martin Davis. *Computability and Unsolvability*. Dover Publications, 1985.

[5] Abraham Adolf Fraenkel, Yehoshua Bar-Hillel, and Azriel Levy. *Foundations of Set Theory*, volume 67. North Holland, 1973.

[6] Galileo Galilei. *Dialogues Concerning Two New Sciences*. Prometheus Books, 1991.

[7] Paul Richard Halmos. *Naive Set Theory*. Springer, 1974.

[8] Andrew Hodges. *Alan Turing: The Enigma*. Random House, 1992.

[9] Jean Van Heijenoort. *From Frege to Gödel: a Source Book in Mathematical Logic, 1879-1931*. Harvard University Press, 1967.