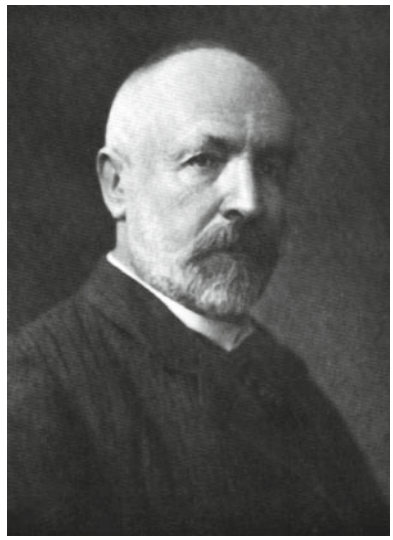
Sets, functions, and the continuum hypothesis

Chapter 19



Georg Cantor

Set theory, founded by Georg Cantor in the second half of the 19th century, has profoundly transformed mathematics. Modern day mathematics is unthinkable without the concept of a set, or as David Hilbert put it: "Nobody will drive us from the paradise (of set theory) that Cantor has created for us."

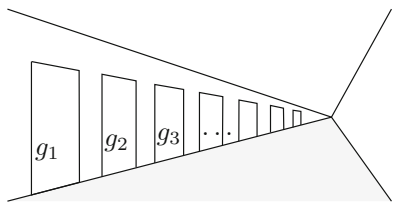
One of Cantor's basic concepts is the notion of the *size* or *cardinality* of a set M, denoted by $|M|$. For finite sets, this presents no difficulties: we just count the number of elements and say that M is an n-set or has size n, if M contains precisely n elements. Thus two finite sets M and N have equal size, $|M| = |N|$, if they contain the same number of elements.

To carry this notion of *equal* size over to infinite sets, we use the following suggestive thought experiment for finite sets. Suppose a number of people board a bus. When will we say that the number of people is the same as the number of available seats? Simple enough, we let all people sit down. If everyone finds a seat, and no seat remains empty, then and only then do the two sets (of the people and of the seats) agree in number. In other words, the two sizes are the same if there is a *bijection* of one set onto the other.

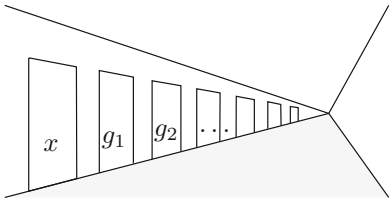
This is then our definition: Two arbitrary sets M and N (finite or infinite) are said to be of *equal size* or *cardinality*, if and only if there exists a bijection from M onto N. Clearly, this notion of equal size is an equivalence relation, and we can thus associate a number, called *cardinal number*, to every class of equal-sized sets. For example, we obtain for finite sets the cardinal numbers $0, 1, 2, \ldots, n, \ldots$ where n stands for the class of n-sets, and, in particular, 0 for the *empty set* $\varnothing$. We further observe the obvious fact that a proper subset of a finite set M invariably has smaller size than M.

The theory becomes very interesting (and highly non-intuitive) when we turn to infinite sets. Consider the set $\mathbb{N} = \{1, 2, 3, \ldots\}$ of natural numbers. We call a set M *countable* if it can be put in one-to-one correspondence with $\mathbb{N}$. In other words, M is countable if we can list the elements of M as $m_1, m_2, m_3, \ldots$. But now a strange phenomenon occurs. Suppose we add to $\mathbb{N}$ a new element x. Then $\mathbb{N} \cup \{x\}$ is still countable, and hence has equal size with $\mathbb{N}$!

This fact is delightfully illustrated by "Hilbert's hotel." Suppose a hotel has countably many rooms, numbered $1, 2, 3, \ldots$ with guest g_i occupying room i; so the hotel is fully booked. Now a new guest x arrives asking for a room, whereupon the hotel manager tells him: Sorry, all rooms are taken. No problem, says the new arrival, just move guest g_1 to room 2, g_2 to room 3, g_3 to room 4, and so on, and I will then take room 1. To the



M. Aigner, G. M. Ziegler, *Proofs from THE BOOK*, https://doi.org/10.1007/978-3-662-57265-8_19

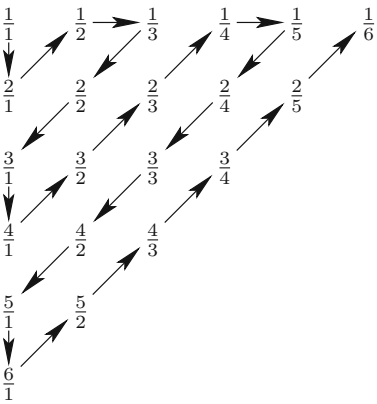


manager's surprise (he is not a mathematician) this works; he can still put up all guests plus the new arrival x!

Now it is clear that he can also put up another guest y, and another one z, and so on. In particular, we note that, in contrast to finite sets, it may well happen that a proper subset of an *infinite* set M has the same size as M. In fact, as we will see, this is a characterization of infinity: A set is infinite if and only if it has the same size as some proper subset.

Let us leave Hilbert's hotel and look at our familiar number sets. The set $\mathbb{Z}$ of integers is again countable, since we may enumerate $\mathbb{Z}$ in the form $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \ldots\}$. It may come more as a surprise that the rationals can be enumerated in a similar way.

Theorem 1. *The set $\mathbb{Q}$ of rational numbers is countable.*



■ **Proof.** By listing the set $\mathbb{Q}^+$ of positive rationals as suggested in the figure in the margin, but leaving out numbers already encountered, we see that $\mathbb{Q}^+$ is countable, and hence so is $\mathbb{Q}$ by listing 0 at the beginning and $-\frac{p}{q}$ right after $\frac{p}{q}$. With this listing

$$\mathbb{Q} = \{0, 1, -1, 2, -2, \tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{3}, -\tfrac{1}{3}, 3, -3, 4, -4, \tfrac{3}{2}, -\tfrac{3}{2}, \ldots\}. \qquad \square$$

Another way to interpret the figure is the following statement:

The union of countably many countable sets M_n is again countable.

Indeed, set $M_n = \{a_{n1}, a_{n2}, a_{n3}, \ldots\}$ and list

$$\bigcup_{n=1}^{\infty} M_n = \{a_{11}, a_{21}, a_{12}, a_{13}, a_{22}, a_{31}, a_{41}, a_{32}, a_{23}, a_{14}, \ldots\}$$

precisely as before.

Let us contemplate Cantor's enumeration of the positive rationals a bit more. Looking at the figure we obtained the sequence

$$\tfrac{1}{1}, \tfrac{2}{1}, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{2}{2}, \tfrac{3}{1}, \tfrac{4}{1}, \tfrac{3}{2}, \tfrac{2}{3}, \tfrac{1}{4}, \tfrac{1}{5}, \tfrac{2}{4}, \tfrac{3}{3}, \tfrac{4}{2}, \tfrac{5}{1}, \ldots$$

and then had to strike out the duplicates such as $\frac{2}{2} = \frac{1}{1}$ or $\frac{2}{4} = \frac{1}{2}$.

But there is a listing that is even more elegant and systematic, and which contains no duplicates — found only quite recently by Neil Calkin and Herbert Wilf. Their new list starts as follows:

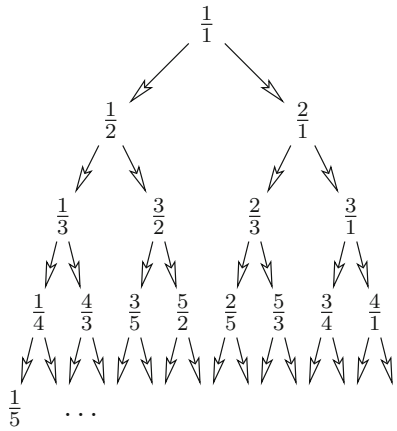
$$\tfrac{1}{1}, \tfrac{1}{2}, \tfrac{2}{1}, \tfrac{1}{3}, \tfrac{3}{2}, \tfrac{2}{3}, \tfrac{3}{1}, \tfrac{1}{4}, \tfrac{4}{3}, \tfrac{3}{5}, \tfrac{5}{2}, \tfrac{2}{5}, \tfrac{5}{3}, \tfrac{3}{4}, \tfrac{4}{1}, \ldots .$$

Here the denominator of the n-th rational number equals the numerator of the $(n+1)$-st number. In other words, the n-th fraction is $b(n)/b(n+1)$, where $\big(b(n)\big)_{n\geq 0}$ is a sequence that starts with

$$(1, 1, 2, 1, 3, 2, 3, 1, 4, 3, 5, 2, 5, 3, 4, 1, 5, \ldots).$$

This sequence has first been studied by a German mathematician, Moritz Abraham Stern, in a paper from 1858, and is has become known as "Stern's diatomic series."

How do we obtain this sequence, and hence the Calkin–Wilf listing of the positive fractions? Consider the infinite binary tree in the margin. We immediately note its recursive rule:



- $\frac{1}{1}$ is on top of the tree, and
- every node $\frac{i}{j}$ has two sons: the left son is $\frac{i}{i+j}$ and the right son is $\frac{i+j}{j}$.

We can easily check the following four properties:

(1) All fractions in the tree are reduced, that is, if $\frac{r}{s}$ appears in the tree, then r and s are relatively prime.

This holds for the top $\frac{1}{1}$, and then we use induction downward. If r and s are relatively prime, then so are r and $r+s$, as well as s and $r+s$.

(2) Every reduced fraction $\frac{r}{s} > 0$ appears in the tree.

We use induction on the sum $r+s$. The smallest value is $r+s=2$, that is $\frac{r}{s}=\frac{1}{1}$, and this appears at the top. If $r>s$, then $\frac{r-s}{s}$ appears in the tree by induction, and so we get $\frac{r}{s}$ as its right son. Similarly, if $r<s$, then $\frac{r}{s-r}$ appears, which has $\frac{r}{s}$ as its left son.

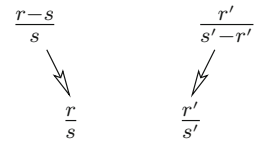
(3) Every reduced fraction appears exactly once.

The argument is similar. If $\frac{r}{s}$ appears more than once, then $r \neq s$, since any node in the tree except the top is of the form $\frac{i}{i+j} < 1$ or $\frac{i+j}{j} > 1$. But if $r > s$ or $r < s$, then we argue by induction as before.

Every positive rational appears therefore exactly once in our tree, and we may write them down listing the numbers level-by-level from left to right. This yields precisely the initial segment shown above.

(4) The denominator of the n-th fraction in our list equals the numerator of the $(n+1)$-st.

This is certainly true for $n=0$, or when the n-th fraction is a left son. Suppose the n-th number $\frac{r}{s}$ is a right son. If $\frac{r}{s}$ is at the right boundary, then $s=1$, and the successor lies at the left boundary and has numerator 1. Finally, if $\frac{r}{s}$ is in the interior, and $\frac{r'}{s'}$ is the next fraction in our sequence, then $\frac{r}{s}$ is the right son of $\frac{r-s}{s}$, $\frac{r'}{s'}$ is the left son of $\frac{r'}{s'-r'}$, and by induction the denominator of $\frac{r-s}{s}$ is the numerator of $\frac{r'}{s'-r'}$, so we get $s=r'$.



Well, this is nice, but there is even more to come. There are two natural questions:

– Does the sequence $\big(b(n)\big)_{n\geq 0}$ have a "meaning"? That is, does $b(n)$ count anything simple?

– Given $\frac{r}{s}$, is there an easy way to determine the successor in the listing?

To answer the first question, we work out that the node $b(n)/b(n+1)$ has the two sons $b(2n+1)/b(2n+2)$ and $b(2n+2)/b(2n+3)$. By the set-up of the tree we obtain the recursions

$$b(2n+1) = b(n) \quad \text{and} \quad b(2n+2) = b(n) + b(n+1). \tag{1}$$

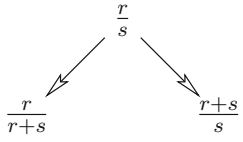
With $b(0) = 1$ the sequence $(b(n))_{n\geq 0}$ is completely determined by (1).

So, is there a "nice" "known" sequence which obeys the same recursion? Yes, there is. We know that any number n can be uniquely written as a sum of distinct powers of 2 — this is the usual binary representation of n. A *hyper-binary* representation of n is a representation of n a sum of powers of 2, where every power 2^k appears at most *twice*. Let $h(n)$ be the number of such representations for n. You are invited to check that the sequence $h(n)$ obeys the recursion (1), and this gives $b(n) = h(n)$ for all n.

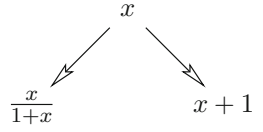
For example, $h(6) = 3$, with the hyper-binary representations
$6 = 4 + 2$
$6 = 4 + 1 + 1$
$6 = 2 + 2 + 1 + 1$.

Incidentally, we have proved a surprising fact: Let $\frac{r}{s}$ be a reduced fraction, there exists precisely one integer n with $r = h(n)$ and $s = h(n+1)$.

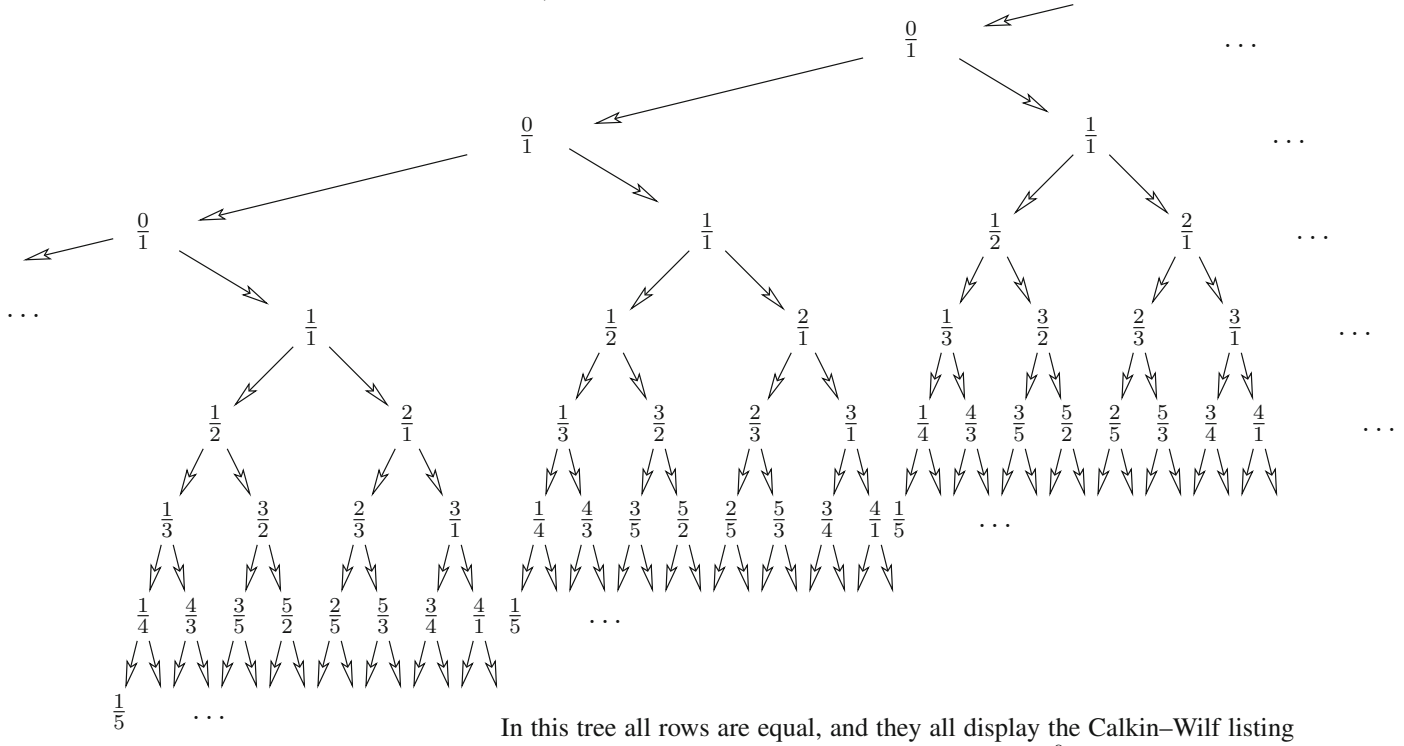
Let us look at the second question. We have in our tree



that is, with $x := \frac{r}{s}$,



We now use this to generate an even larger infinite binary tree (without a root) as follows:



In this tree all rows are equal, and they all display the Calkin–Wilf listing of the positive rationals (starting with an additional $\frac{0}{1}$).

So how does one get from one rational to the next? To answer this, we first record that for every rational x its right son is $x+1$, the right grand-son is $x+2$, so the k-fold right son is $x+k$. Similarly, the left son of x is $\frac{x}{1+x}$, whose left son is $\frac{x}{1+2x}$, and so on: The k-fold left son of x is $\frac{x}{1+kx}$.
Now to find how to get from $\frac{r}{s} = x$ to the "next" rational $f(x)$ in the listing, we have to analyze the situation depicted in the margin. In fact, if we consider any nonnegative rational number x in our infinite binary tree, then it is the k-fold right son of the left son of some rational $y \geq 0$ (for some $k \geq 0$), while $f(x)$ is given as the k-fold left son of the right son of the same y. Thus with the formulas for k-fold left sons and k-fold right sons, we get

$$x \;=\; \frac{y}{1+y} + k,$$

as claimed in the figure in the margin. Here $k = \lfloor x \rfloor$ is the integral part of x, while $\frac{y}{1+y} = \{x\}$ is the fractional part. And from this we obtain

$$f(x) = \frac{y+1}{1+k(y+1)} = \frac{1}{\frac{1}{y+1}+k} = \frac{1}{k+1-\frac{y}{y+1}} = \frac{1}{\lfloor x \rfloor + 1 - \{x\}}\,.$$

Thus we have obtained a beautiful formula for the successor $f(x)$ of x, first found by Moshe Newman:

> *The function*
>
> $$x \;\longmapsto\; f(x) \;=\; \frac{1}{\lfloor x \rfloor + 1 - \{x\}}$$
>
> *generates the Calkin–Wilf sequence*
>
> $$\tfrac{1}{1} \mapsto \tfrac{1}{2} \mapsto \tfrac{2}{1} \mapsto \tfrac{1}{3} \mapsto \tfrac{3}{2} \mapsto \tfrac{2}{3} \mapsto \tfrac{3}{1} \mapsto \tfrac{1}{4} \mapsto \tfrac{4}{3} \mapsto \cdots$$
>
> *which contains every positive rational number exactly once.*

The Calkin–Wilf–Newman way to enumerate the positive rationals has a number of additional remarkable properties. For example, one may ask for a fast way to determine the n-th fraction in the sequence, say for $n = 10^6$. Here it is:

> To find the n-th fraction in the Calkin–Wilf sequence, express n as a binary number $n = (b_k b_{k-1} \ldots b_1 b_0)_2$, and then follow the path in the Calkin–Wilf tree that is determined by its digits, starting at $\frac{s}{t} = \frac{0}{1}$.
> Here $b_i = 1$ means "take the right son," that is, "add the denominator to the numerator," while $b_i = 0$ means "take the left son," that is, "add the numerator to the denominator."

The figure in the margin shows the resulting path for $n = 25 = (11001)_2$: So the 25th number in the Calkin–Wilf sequence is $\frac{7}{5}$. The reader could easily work out a similar scheme that computes for a given fraction $\frac{s}{t}$ (the binary representation of) its position n in the Calkin–Wilf sequence.

Let us move on to the real numbers $\mathbb{R}$. Are they still countable? No, they are not, and the means by which this is shown — Cantor's *diagonalization method* — is not only of fundamental importance for all of set theory, but certainly belongs into The Book as a rare stroke of genius.

Theorem 2. *The set $\mathbb{R}$ of real numbers is **not** countable.*

■ **Proof.** Any subset N of a countable set $M = \{m_1, m_2, m_3, \ldots\}$ is *at most countable* (that is, finite or countable). In fact, just list the elements of N as they appear in M. Accordingly, if we can find a subset of $\mathbb{R}$ which is not countable, then a fortiori $\mathbb{R}$ cannot be countable. The subset M of $\mathbb{R}$ we want to look at is the interval $(0, 1]$ of all positive real numbers r with $0 < r \leq 1$. Suppose, to the contrary, that M is countable, and let $M = \{r_1, r_2, r_3, \ldots\}$ be a listing of M. We write r_n as its unique *infinite* decimal expansion without an infinite sequence of zeros at the end:

$$r_n = 0.a_{n1}a_{n2}a_{n3}...$$

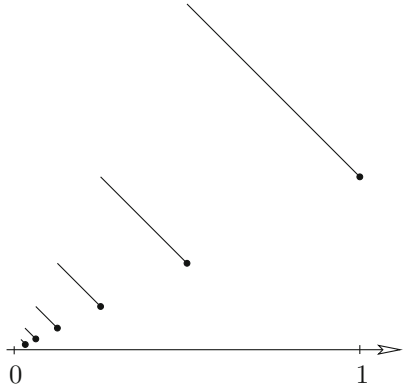
where $a_{ni} \in \{0, 1, \ldots, 9\}$ for all n and i. For example, $0.7 = 0.6999...$ Consider now the doubly infinite array

$$\begin{array}{rcl} r_1 & = & 0.a_{11}a_{12}a_{13}... \\ r_2 & = & 0.a_{21}a_{22}a_{23}... \\ & \vdots & \\ r_n & = & 0.a_{n1}a_{n2}a_{n3}... \\ & \vdots & \end{array}$$

For every n, let b_n be the least element of $\{1, 2\}$ that is different from a_{nn}. Then $b = 0.b_1b_2b_3...b_n...$ is a real number in our set M and hence must have an index, say $b = r_k$. But this cannot be, since b_k is different from a_{kk}. And this is the whole proof! □

Let us stay with the real numbers for a moment. We note that all four types of intervals $(0, 1), (0, 1], [0, 1)$ and $[0, 1]$ have the same size. As an example, we verify that $(0, 1]$ and $(0, 1)$ have equal cardinality. The map $f : (0, 1] \longrightarrow (0, 1), x \longmapsto y$ defined by

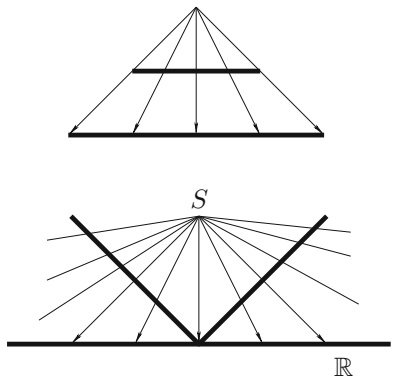
$$y \coloneqq \begin{cases} \frac{3}{2} - x & \text{for} \quad \frac{1}{2} < x \leq 1, \\ \frac{3}{4} - x & \text{for} \quad \frac{1}{4} < x \leq \frac{1}{2}, \\ \frac{3}{8} - x & \text{for} \quad \frac{1}{8} < x \leq \frac{1}{4}, \\ \quad \vdots & \end{cases}$$



A bijective $f : (0, 1] \longrightarrow (0, 1)$

does the job. Indeed, the map is bijective, since the range of y in the first line is $\frac{1}{2} \leq y < 1$, in the second line $\frac{1}{4} \leq y < \frac{1}{2}$, in the third line $\frac{1}{8} \leq y < \frac{1}{4}$, and so on.

Next we find that *any* two intervals (of finite length > 0) have equal size by considering the central projection as in the figure. Even more is true: Every interval (of length > 0) has the same size as the whole real line $\mathbb{R}$. To see this, look at the bent open interval $(0, 1)$ and project it onto $\mathbb{R}$ from the center S.



So, in conclusion, any open, half-open, closed (finite or infinite) interval of length > 0 has the same size, and we denote this size by c, where c stands for *continuum* (a name sometimes used for the interval [0,1]).

That finite and infinite intervals have the same size may come expected on second thought, but here is a fact that is downright counter-intuitive.

Theorem 3. *The set $\mathbb{R}^2$ of all ordered pairs of real numbers (that is, the real plane) has the same size as $\mathbb{R}$.*

The theorem is due to Cantor 1878, as is the idea to merge the decimal expansions of two reals into one. The variant of Cantor's method that we are going to present is again from The Book. Abraham Fraenkel attributes the trick, which directly yields a bijection, to Julius König.

■ **Proof.** It suffices to prove that the set of all pairs (x, y), $0 < x, y \leq 1$, can be mapped bijectively onto $(0, 1]$. Consider the pair (x, y) and write x, y in their unique non-terminating decimal expansion as in the following example:

$$\begin{array}{rcrrrrrr} x & = & 0.3 & 01 & 2 & 007 & 08 & \ldots \\ y & = & 0.009 & 2 & 05 & 1 & 0008 & \ldots \end{array}$$

Note that we have separated the digits of x and y into groups by always going to the next nonzero digit, inclusive. Now we associate to (x, y) the number $z \in (0, 1]$ by writing down the first x-group, after that the first y-group, then the second x-group, and so on. Thus, in our example, we obtain

$$z = 0.3\,009\,01\,2\,2\,05\,007\,1\,08\,0008\ \ldots$$

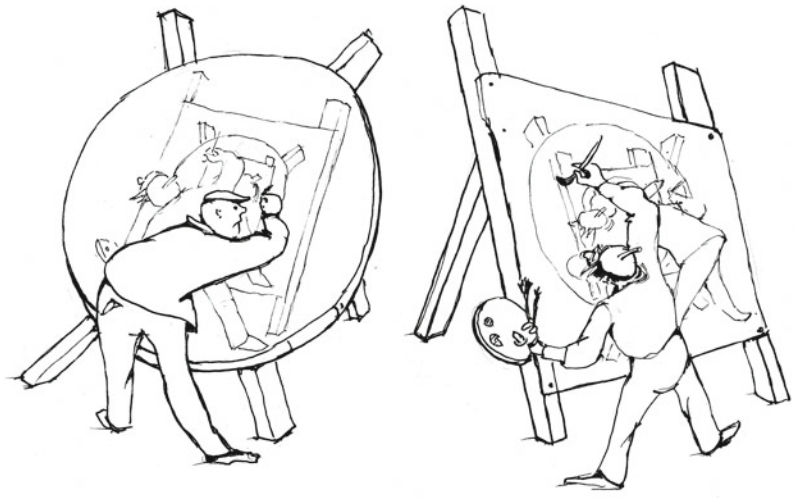
Since neither x nor y exhibits only zeros from a certain point on, we find that the expression for z is again a non-terminating decimal expansion. Conversely, from the expansion of z we can immediately read off the preimage (x, y), and the map is bijective — end of proof. □

As $(x, y) \longmapsto x + iy$ is a bijection from $\mathbb{R}^2$ onto the complex numbers $\mathbb{C}$, we conclude that $|\mathbb{C}| = |\mathbb{R}| = c$. Why is the result $|\mathbb{R}^2| = |\mathbb{R}|$ so unexpected? Because it goes against our intuition of *dimension*. It says that the 2-dimensional plane $\mathbb{R}^2$ (and, in general, by induction, the n-dimensional space $\mathbb{R}^n$) can be mapped bijectively onto the 1-dimensional line $\mathbb{R}$. Thus dimension is not generally preserved by bijective maps. If, however, we require the map and its inverse to be continuous, then the dimension is preserved, as was first shown by Luitzen Brouwer.

Let us go a little further. So far, we have the notion of equal size. When will we say that M is at most as large as N? Mappings provide again the key. We say that the cardinal number $\mathfrak{m}$ is *less than or equal to* $\mathfrak{n}$, if for sets M and N with $|M| = \mathfrak{m}$, $|N| = \mathfrak{n}$, there exists an *injection* from M into N. Clearly, the relation $\mathfrak{m} \leq \mathfrak{n}$ is independent of the representative sets M and N chosen. For finite sets this corresponds again to our intuitive notion: An m-set is at most as large as an n-set if and only if $m \leq n$.

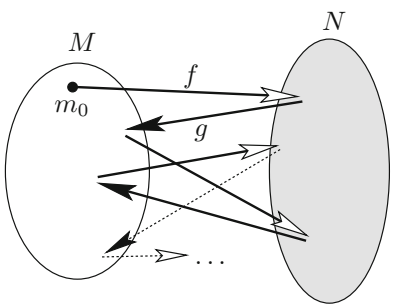
Now we are faced with a basic problem. We would certainly like to have that the usual laws concerning inequalities also hold for cardinal numbers. But is this true for infinite cardinals? In particular, is it true that $\mathfrak{m} \leq \mathfrak{n}$, $\mathfrak{n} \leq \mathfrak{m}$ imply $\mathfrak{m} = \mathfrak{n}$?

The affirmative answer to this question is provided by the famous Cantor–Bernstein theorem, which Cantor announced in 1883. The first complete proof was presented by Felix Bernstein in Cantor's seminar in 1897. Further proofs were given by Richard Dedekind, Ernst Zermelo, and others. Our proof is due to Julius König (1906).



"Cantor and Bernstein painting"

Theorem 4. *If each of two sets M and N can be mapped injectively into the other, then there is a bijection from M to N, that is, $|M| = |N|$.*



■ **Proof.** We may certainly assume that M and N are disjoint — if not, then we just replace N by a new copy.

Now f and g map back and forth between the elements of M and those of N. One way to bring this potentially confusing situation into perfect clarity and order is to align $M \cup N$ into chains of elements: Take an arbitrary element $m_0 \in M$, say, and from this generate a chain of elements by applying f, then g, then f again, then g, and so on. The chain may close up (this is Case 1) if we reach m_0 again in this process, or it may continue with distinct elements indefinitely. (The first "duplicate" in the chain cannot be an element different from m_0, by injectivity.)

If the chain continues indefinitely, then we try to follow it backwards: From m_0 to $g^{-1}(m_0)$ if m_0 is in the image of g, then to $f^{-1}(g^{-1}(m_0))$ if $g^{-1}(m_0)$ is in the image of f, and so on. Three more cases may arise here: The process of following the chain backwards may go on indefinitely (Case 2), it may stop in an element of M that does not lie in the image of g (Case 3), or it may stop in an element of N that does not lie in the image of f (Case 4).

Thus $M \cup N$ splits perfectly into four types of chains, whose elements we may label in such a way that a bijection is simply given by putting $F : m_i \longmapsto n_i$. We verify this in the four cases separately:

Case 1. Finite cycles on $2k + 2$ distinct elements ($k \geq 0$)

$$m_0 \overset{f}{\twoheadrightarrow} n_0 \overset{g}{\rightarrow} m_1 \overset{f}{\twoheadrightarrow} \cdots \quad m_k \overset{f}{\twoheadrightarrow} n_k \overset{g}{\rightarrow} m_0$$

Case 2. Two-way infinite chains of distinct elements

$$\cdots \rightarrow m_0 \overset{f}{\twoheadrightarrow} n_0 \overset{g}{\rightarrow} m_1 \overset{f}{\twoheadrightarrow} n_1 \overset{g}{\rightarrow} m_2 \overset{f}{\twoheadrightarrow} \cdots$$

Case 3. The one-way infinite chains of distinct elements that start at the elements $m_0 \in M \backslash g(N)$

$$m_0 \overset{f}{\twoheadrightarrow} n_0 \overset{g}{\rightarrow} m_1 \overset{f}{\twoheadrightarrow} n_1 \overset{g}{\rightarrow} m_2 \overset{f}{\twoheadrightarrow} \cdots$$

Case 4. The one-way infinite chains of distinct elements that start at the elements $n_0 \in N \backslash f(M)$

$$n_0 \overset{g}{\rightarrow} m_0 \overset{f}{\twoheadrightarrow} n_1 \overset{g}{\rightarrow} m_1 \overset{f}{\twoheadrightarrow} \cdots \quad \square$$

What about the other relations governing inequalities? As usual, we set $\mathfrak{m} < \mathfrak{n}$ if $\mathfrak{m} \leq \mathfrak{n}$, but $\mathfrak{m} \neq \mathfrak{n}$. We have just seen that for any two cardinals $\mathfrak{m}$ and $\mathfrak{n}$ at most one of the three possibilities

$$\mathfrak{m} < \mathfrak{n}, \ \mathfrak{m} = \mathfrak{n}, \ \mathfrak{m} > \mathfrak{n}$$

holds, and it follows from the theory of cardinal numbers that, in fact, precisely one relation is true. (See the appendix to this chapter, Proposition 2.)

Furthermore, the Cantor–Bernstein Theorem tells us that the relation $<$ is transitive, that is, $\mathfrak{m} < \mathfrak{n}$ and $\mathfrak{n} < \mathfrak{p}$ imply $\mathfrak{m} < \mathfrak{p}$. Thus the cardinalities are arranged in linear order starting with the finite cardinals $0, 1, 2, 3, \ldots$. Invoking the usual Zermelo–Fraenkel axiom system, we easily find that any infinite set M contains a countable subset. In fact, M contains an element, say m_1. The set $M \setminus \{m_1\}$ is not empty (since it is infinite) and hence contains an element m_2. Considering $M \setminus \{m_1, m_2\}$ we infer the existence of m_3, and so on. So, the size of a countable set is the *smallest infinite* cardinal, usually denoted by $\aleph_0$ (pronounced "aleph zero").

"The smallest infinite cardinal"

As a corollary to $\aleph_0 \leq \mathfrak{m}$ for any infinite cardinal $\mathfrak{m}$, we can immediately prove "Hilbert's hotel" for any infinite cardinal number $\mathfrak{m}$, that is, we have $|M \cup \{x\}| = |M|$ for any infinite set M. Indeed, M contains a subset $N = \{m_1, m_2, m_3, \ldots\}$. Now map x onto m_1, m_1 onto m_2, and so on, keeping the elements of $M \backslash N$ fixed. This gives the desired bijection.

With this we have also proved a result announced earlier: *Every infinite set has the same size as some proper subset.*

As another consequence of the Cantor–Bernstein theorem we may prove that the set $\mathcal{P}(\mathbb{N})$ of all subsets of $\mathbb{N}$ has cardinality c. As noted above, it suffices to show that $|\mathcal{P}(\mathbb{N}) \backslash \{\varnothing\}| = |(0,1]|$. An example of an injective map is

$$\begin{aligned} f:\ \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\} &\longrightarrow (0,1], \\ A &\longmapsto \sum_{i \in A} 10^{-i}, \end{aligned}$$

while

$$\begin{aligned} g:\ (0,1] &\longrightarrow \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\}, \\ 0.b_1 b_2 b_3 \ldots &\longmapsto \{b_i 10^i : i \in \mathbb{N}\} \end{aligned}$$

defines an injection in the other direction.

Up to now we know the cardinal numbers $0, 1, 2, \ldots, \aleph_0$, and further that the cardinality c of $\mathbb{R}$ is bigger than $\aleph_0$. The passage from $\mathbb{Q}$ with $|\mathbb{Q}| = \aleph_0$ to $\mathbb{R}$ with $|\mathbb{R}| = c$ immediately suggests the next question:

Is $c = |\mathbb{R}|$ the next infinite cardinal number after $\aleph_0$?

Now, of course, we have the problem whether there *is* a next larger cardinal number, or in other words, whether $\aleph_1$ has a meaning at all. It does — the proof for this is outlined in the appendix to this chapter.

The statement $c = \aleph_1$ became known as the *continuum hypothesis*. The question whether the continuum hypothesis is true presented for many decades one of the supreme challenges in all of mathematics. The answer, finally given by Kurt Gödel and Paul Cohen, takes us to the limit of logical thought. They showed that the statement $c = \aleph_1$ is *independent* of the Zermelo–Fraenkel axiom system, in the same way as the parallel axiom is independent of the other axioms of Euclidian geometry. There are models where $c = \aleph_1$ holds, and there are other models of set theory where $c \neq \aleph_1$ holds.

In the light of this fact it is quite interesting to ask whether there are other conditions (from analysis, say) which are equivalent to the continuum hypothesis. Indeed, it is natural to ask for an analysis example, since historically the first substantial applications of Cantor's set theory occurred in analysis, specifically in complex function theory. In the following we want to present one such instance and its extremely elegant and simple solution by Paul Erdős. In 1962 John E. Wetzel, a young instructor at the University of Illinois, asked the following question:

> *Let $\{f_\alpha\}$ be a family of pairwise distinct analytic functions on the complex numbers such that for each $z \in \mathbb{C}$ the set of values $\{f_\alpha(z)\}$ is at most countable (that is, it is either finite or countable); let us call this property (P_0).*
> *Does it then follow that the family itself is at most countable?*

Very shortly afterwards Erdős showed that, surprisingly, the answer depends on the continuum hypothesis.

Theorem 5. *If $c > \aleph_1$, then every family $\{f_\alpha\}$ satisfying (P_0) is countable. If, on the other hand, $c = \aleph_1$, then there exists some family $\{f_\alpha\}$ with property (P_0) which has size c.*

For the proof we need some basic facts on cardinal and ordinal numbers. For readers who are unfamiliar with these concepts, this chapter has an appendix where all the necessary results are collected.

■ **Proof.** Assume first $c > \aleph_1$. We shall show that for any family $\{f_\alpha\}$ of size $\aleph_1$ of analytic functions there exists a complex number z_0 such that *all* $\aleph_1$ values $f_\alpha(z_0)$ are distinct. Consequently, if a family of functions satisfies (P_0), then it must be countable.

To see this, we make use of our knowledge of ordinal numbers. First, we well-order the family $\{f_\alpha\}$ according to the initial ordinal number ω_1 of $\aleph_1$. This means by Proposition 1 of the appendix that the index set runs through all ordinal numbers α which are smaller than ω_1. Next we show that the set of pairs (α, β), $\alpha < \beta < \omega_1$, has size $\aleph_1$. Since any $\beta < \omega_1$ is a countable ordinal, the set of pairs (α, β), $\alpha < \beta$, is countable for every fixed β. Taking the union over all $\aleph_1$-many β, we find from Proposition 6 of the appendix that the set of all pairs (α, β), $\alpha < \beta$, has size $\aleph_1$.

Consider now for any pair $\alpha < \beta$ the set

$$S(\alpha, \beta) \;=\; \{z \in \mathbb{C} : f_\alpha(z) = f_\beta(z)\}.$$

We claim that each set $S(\alpha, \beta)$ is countable. To verify this, consider the disks C_k of radius $k = 1, 2, 3, \ldots$ around the origin in the complex plane. If f_α and f_β agree on infinitely many points in some C_k, then f_α and f_β are identical by a well-known result on analytic functions. Hence f_α and f_β agree only in finitely many points in each C_k, and hence in at most countably many points altogether. Now we set

$$S \;\coloneqq\; \bigcup_{\alpha<\beta} S(\alpha, \beta).$$

Again by Proposition 6, we find that S has size $\aleph_1$, as each set $S(\alpha, \beta)$ is countable. And here is the punch line: Because, as we know, $\mathbb{C}$ has size c, and c is larger than $\aleph_1$ by assumption, there exists a complex number z_0 not in S, and for this z_0 all $\aleph_1$ values $f_\alpha(z_0)$ are distinct.

Next we assume $c = \aleph_1$. Consider the set $D \subseteq \mathbb{C}$ of complex numbers $p + iq$ with rational real and imaginary part. Since for each p the set $\{p + iq : q \in \mathbb{Q}\}$ is countable, we find that D is countable. Furthermore, D is a *dense* set in $\mathbb{C}$: Every open disk in the complex plane contains some point of D. Let $\{z_\alpha : 0 \leq \alpha < \omega_1\}$ be a well-ordering of $\mathbb{C}$. We shall now construct a family $\{f_\beta : 0 \leq \beta < \omega_1\}$ of $\aleph_1$-many distinct analytic functions such that

$$f_\beta(z_\alpha) \in D \text{ whenever } \alpha < \beta. \tag{1}$$

Any such family satisfies the condition (P_0). Indeed, each point $z \in \mathbb{C}$ has some index, say $z = z_\alpha$. Now, for all $\beta > \alpha$, the values $\{f_\beta(z_\alpha)\}$ lie in the *countable* set D. Since α is a countable ordinal number, the functions f_β with $\beta \leq \alpha$ will contribute at most countably further values $f_\beta(z_\alpha)$, so that the set of all values $\{f_\beta(z_\alpha)\}$ is likewise at most countable. Hence, if we can construct a family $\{f_\beta\}$ satisfying (1), then the second part of the theorem is proved.

The construction of $\{f_\beta\}$ is by transfinite induction. For f_0 we may take any analytic function, for example $f_0 =$ constant. Suppose f_β has already been constructed for all $\beta < \gamma$. Since γ is a countable ordinal, we may reorder $\{f_\beta : 0 \leq \beta < \gamma\}$ into a sequence $g_1, g_2, g_3, \ldots$. The same reordering of $\{z_\alpha : 0 \leq \alpha < \gamma\}$ yields a sequence $w_1, w_2, w_3, \ldots$. We shall now construct a function f_γ satisfying for each n the conditions

$$f_\gamma(w_n) \in D \qquad \text{and} \qquad f_\gamma(w_n) \neq g_n(w_n). \tag{2}$$

The second condition will ensure that all functions f_γ $(0 \leq \gamma < \omega_1)$ are distinct, and the first condition is just (1), implying (P_0) by our previous argument. Notice that the condition $f_\gamma(w_n) \neq g_n(w_n)$ is once more a diagonalization argument.

To construct f_γ, we write

$$\begin{aligned} f_\gamma(z) \;\coloneqq\; & \varepsilon_0 + \varepsilon_1(z - w_1) + \varepsilon_2(z - w_1)(z - w_2) \\ & + \varepsilon_3(z - w_1)(z - w_2)(z - w_3) + \cdots . \end{aligned}$$

If γ is a finite ordinal, then f_γ is a polynomial and hence analytic, and we can certainly choose numbers ε_i such that (2) is satisfied. Now suppose γ is a countable ordinal, then

$$f_\gamma(z) \;=\; \sum_{n=0}^{\infty} \varepsilon_n (z - w_1) \cdots (z - w_n). \tag{3}$$

Note that the values of ε_m $(m \geq n)$ have no influence on the value $f_\gamma(w_n)$, hence we may choose the ε_n step by step. If the sequence (ε_n) converges to 0 sufficiently fast, then (3) defines an analytic function. Finally, since D is a dense set, we may choose this sequence (ε_n) so that f_γ meets the requirements of (2), and the proof is complete. □

Appendix: On cardinal and ordinal numbers

Let us first discuss the question whether to each cardinal number there exists a next larger one. As a start we show that to every cardinal number $\mathfrak{m}$ there always is a cardinal number $\mathfrak{n}$ larger than $\mathfrak{m}$. To do this we employ again a version of Cantor's diagonalization method.

Let M be a set, then we claim that the set $\mathcal{P}(M)$ *of all subsets* of M has larger size than M. By letting $m \in M$ correspond to $\{m\} \in \mathcal{P}(M)$, we see that M can be mapped bijectively onto a subset of $\mathcal{P}(M)$, which implies $|M| \leq |\mathcal{P}(M)|$ by definition. It remains to show that $\mathcal{P}(M)$ *can not* be mapped bijectively onto a subset of M. Suppose, on the contrary, $\varphi : N \longrightarrow \mathcal{P}(M)$ is a bijection of $N \subseteq M$ onto $\mathcal{P}(M)$. Consider the subset $U \subseteq N$ of all elements of N which are *not* contained in their image under φ, that is, $U = \{m \in N : m \notin \varphi(m)\}$. Since φ is a bijection, there exists $u \in N$ with $\varphi(u) = U$. Now, either $u \in U$ or $u \notin U$, but both alternatives are impossible! Indeed, if $u \in U$, then $u \notin \varphi(u) = U$ by the definition of U, and if $u \notin U = \varphi(u)$, then $u \in U$, contradiction.

Most likely, the reader has seen this argument before. It is the old barber riddle: "A barber is the man who shaves all men who do not shave themselves. Does the barber shave himself?"

To get further in the theory we introduce another great concept of Cantor's, ordered sets and ordinal numbers. A set M is *ordered* by $<$ if the relation $<$ is transitive, and if for any two distinct elements a and b of M we either have $a < b$ or $b < a$. For example, we can order $\mathbb{N}$ in the usual way according to magnitude, $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$, but, of course, we can also order $\mathbb{N}$ the other way round, $\mathbb{N} = \{\ldots, 4, 3, 2, 1\}$, or $\mathbb{N} = \{1, 3, 5, \ldots, 2, 4, 6, \ldots\}$ by listing first the odd numbers and then the even numbers.

Here is the seminal concept. An ordered set M is called *well-ordered* if every nonempty subset of M has a first element. Thus the first and third orderings of $\mathbb{N}$ above are well-orderings, but not the second ordering. The fundamental *well-ordering theorem*, implied by the axioms (including the axiom of choice), now states that *every* set M admits a well-ordering. From now on, we only consider sets endowed with a well-ordering.

Let us say that two well-ordered sets M and N are *similar* (or of the *same order-type*) if there exists a bijection φ from M on N which respects the ordering, that is, $m <_M n$ implies $\varphi(m) <_N \varphi(n)$. Note that any ordered set which is similar to a well-ordered set is itself well-ordered.

Similarity is obviously an equivalence relation, and we can thus speak of an *ordinal number* α belonging to a class of similar sets. For finite sets, any two orderings are similar well-orderings, and we use again the ordinal number n for the class of n-sets. Note that, by definition, two similar sets have the same cardinality. Hence it makes sense to speak of the *cardinality* $|\alpha|$ of an ordinal number α. Note further that any subset of a well-ordered set is also well-ordered under the induced ordering.

As we did for cardinal numbers, we now compare ordinal numbers. Let M be a well-ordered set, $m \in M$, then $M_m = \{x \in M : x < m\}$ is called the (*initial*) *segment* of M determined by m; N is a segment of M if $N = M_m$

"A legend talks about St. Augustin who, walking along the seashore and contemplating infinity, saw a child trying to empty the ocean with a small shell…"

The well-ordered sets $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N} = \{1, 3, 5, \ldots, 2, 4, 6, \ldots\}$ are *not* similar: the first ordering has only one element without an immediate predecessor, while the second one has two.

The ordinal number of $\{1, 2, 3, \ldots\}$ is smaller than the ordinal number of $\{1, 3, 5, \ldots, 2, 4, 6, \ldots\}$.

for some m. Thus, in particular, M_m is the empty set when m is the first element of M. Now let μ and ν be the ordinal numbers of the well-ordered sets M and N. We say that μ is *smaller* than ν, $\mu < \nu$, if M is similar to a segment of N. Again, we have the transitive law that $\mu < \nu$, $\nu < \pi$ implies $\mu < \pi$, since under a similarity mapping a segment is mapped onto a segment.

Clearly, for finite sets, $m < n$ corresponds to the usual meaning. Let us denote by ω the ordinal number of $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ ordered according to magnitude. By considering the segment $\mathbb{N}_{n+1}$ we find $n < \omega$ for any finite n. Next we see that $\omega \leq \alpha$ holds for any infinite ordinal number α. Indeed, if the infinite well-ordered set M has ordinal number α, then M contains a first element m_1, the set $M\backslash\{m_1\}$ contains a first element m_2, $M\backslash\{m_1, m_2\}$ contains a first element m_3. Continuing in this way, we produce the sequence $m_1 < m_2 < m_3 < \cdots$ in M. If $M = \{m_1, m_2, m_3, \ldots\}$, then M is similar to $\mathbb{N}$, and hence $\alpha = \omega$. If, on the other hand, $M\backslash\{m_1, m_2, \ldots\}$ is nonempty, then it contains a first element m, and we conclude that $\mathbb{N}$ is similar to the segment M_m, that is, $\omega < \alpha$ by definition.

We now state (without the proofs, which are not difficult) three basic results on ordinal numbers. The first says that any ordinal number μ has a "standard" representative well-ordered set W_μ.

Proposition 1. *Let μ be an ordinal number and denote by W_μ the set of ordinal numbers smaller than μ. Then the following holds:*

(i) *The elements of W_μ are pairwise comparable.*

(ii) *If we order W_μ according to magnitude, then W_μ is well-ordered and has ordinal number μ.*

Proposition 2. *Any two ordinal numbers μ and ν satisfy precisely one of the relations $\mu < \nu$, $\mu = \nu$, or $\mu > \nu$.*

Proposition 3. *Every set of ordinal numbers (ordered according to magnitude) is well-ordered.*

After this excursion to ordinal numbers we come back to cardinal numbers. Let $\mathfrak{m}$ be a cardinal number, and denote by $O_\mathfrak{m}$ the set of all ordinal numbers μ with $|\mu| = \mathfrak{m}$. By Proposition 3 there is a *smallest* ordinal number $\omega_\mathfrak{m}$ in $O_\mathfrak{m}$, which we call the *initial ordinal number* of $\mathfrak{m}$. As an example, ω is the initial ordinal number of $\aleph_0$.

With these preparations we can now prove a basic result for this chapter.

Proposition 4. *For every cardinal number $\mathfrak{m}$ there is a definite next larger cardinal number.*

■ **Proof.** We already know that there is some larger cardinal number $\mathfrak{n}$. Consider now the set $\mathcal{K}$ of all cardinal numbers larger than $\mathfrak{m}$ and at most as large as $\mathfrak{n}$. We associate to each $\mathfrak{p} \in \mathcal{K}$ its initial ordinal number $\omega_\mathfrak{p}$. Among these initial numbers there is a smallest (Proposition 3), and the corresponding cardinal number is then the smallest in $\mathcal{K}$, and thus is the desired next larger cardinal number to $\mathfrak{m}$. □

Proposition 5. *Let the infinite set M have cardinality $\mathfrak{m}$, and let M be well-ordered according to the initial ordinal number $\omega_{\mathfrak{m}}$. Then M has no last element.*

■ **Proof.** Indeed, if M had a last element m, then the segment M_m would have an ordinal number $\mu < \omega_{\mathfrak{m}}$ with $|\mu| = \mathfrak{m}$, contradicting the definition of $\omega_{\mathfrak{m}}$. □

What we finally need is a considerable strenghthening of the result that the union of countably many countable sets is again countable. In the following result we consider *arbitrary* families of countable sets.

Proposition 6. *Suppose $\{A_\alpha\}$ is a family of size $\mathfrak{m}$ of countable sets A_α, where $\mathfrak{m}$ is an infinite cardinal. Then the union $\bigcup_\alpha A_\alpha$ has size at most $\mathfrak{m}$.*

■ **Proof.** We may assume that the sets A_α are pairwise disjoint, since this can only increase the size of the union. Let M with $|M| = \mathfrak{m}$ be the index set, and well-order it according to the initial ordinal number $\omega_{\mathfrak{m}}$. We now replace each $\alpha \in M$ by a countable set $B_\alpha = \{b_{\alpha 1} = \alpha, b_{\alpha 2}, b_{\alpha 3}, \ldots\}$, ordered according to ω, and call the new set $\widetilde{M}$. Then $\widetilde{M}$ is again well-ordered by setting $b_{\alpha i} < b_{\beta j}$ for $\alpha < \beta$ and $b_{\alpha i} < b_{\alpha j}$ for $i < j$. Let $\widetilde{\mu}$ be the ordinal number of $\widetilde{M}$. Since M is a subset of $\widetilde{M}$, we have $\mu \leq \widetilde{\mu}$ by an earlier argument. If $\mu = \widetilde{\mu}$, then M is similar to $\widetilde{M}$, and if $\mu < \widetilde{\mu}$, then M is similar to a segment of $\widetilde{M}$. Now, since the ordering $\omega_{\mathfrak{m}}$ of M has no last element (Proposition 5), we see that M is in both cases similar to the union of countable sets B_β, and hence of the same cardinality.

The rest is easy. Let $\varphi : \bigcup B_\beta \longrightarrow M$ be a bijection, and suppose that $\varphi(B_\beta) = \{\alpha_1, \alpha_2, \alpha_3, \ldots\}$. Replace each α_i by A_{α_i} and consider the union $\bigcup A_{\alpha_i}$. Since $\bigcup A_{\alpha_i}$ is the union of *countably* many countable sets (and hence countable), we see that B_β has the same size as $\bigcup A_{\alpha_i}$. In other words, there is a bijection from B_β to $\bigcup A_{\alpha_i}$ for all β, and hence a bijection ψ from $\bigcup B_\beta$ to $\bigcup A_\alpha$. But now $\psi\varphi^{-1}$ gives the desired bijection from M to $\bigcup A_\alpha$, and thus $|\bigcup A_\alpha| = \mathfrak{m}$. □

References

[1] L. E. J. BROUWER: *Beweis der Invarianz der Dimensionszahl,* Math. Annalen **70** (1911), 161-165.

[2] N. CALKIN & H. WILF: *Recounting the rationals,* Amer. Math. Monthly **107** (2000), 360-363.

[3] G. CANTOR: *Ein Beitrag zur Mannigfaltigkeitslehre,* Journal für die reine und angewandte Mathematik **84** (1878), 242-258.

[4] P. COHEN: *Set Theory and the Continuum Hypothesis,* W. A. Benjamin, New York 1966.

[5] P. ERDŐS: *An interpolation problem associated with the continuum hypothesis,* Michigan Math. J. **11** (1964), 9-10.

[6] E. KAMKE: *Theory of Sets,* Dover Books 1950.

[7] M. A. STERN: *Ueber eine zahlentheoretische Funktion,* Journal für die reine und angewandte Mathematik **55** (1858), 193-220.

"Infinitely many more cardinals"