Contents

Introduction

In the "Mathematical Reality Paradox" article, we saw how the non Euclidean Geometries provoked the concepts of Models and Logic. The Models today of course are nothing more than Structures that satisfy some axioms. More importantly, the Structures themselves are nothing more than Set Constructions. Sets are exactly the abstractions that can be constructed into more complicated sets, so the Models are simply Sets.

Neither Set Theory nor Logic came directly from the new visions of non Euclidean Geometries, that is from these new visions of Mathematical Reality. I'm sure that these new visions impregnated the floating actual problems that mathematicians faced. But to be historically faithful, we should approach both Set Theory and Logic in lines that ignore this subtle impregnation. Luckily, both fields have very elegant and historically faithful approaches. For Logic it is the problem of Aristotelian Formal Logic, while for Set Theory, it is the problem of irrationality.

For Set Theory, the historical approach of course must emphasize the singular importance of Cantor, the discoverer of Sets. This requires a bit of a lamentation on two subjects. The first is the "drop everything" syndrome.

Real creation comes from becoming a medium of the heavens. We can decide to dig deeper and deeper, but only if concrete puzzles guide us to dig. Today, due to the decline of science, in the second half of the twentieth century, we might get the false impression that willful perseverance leads to great results. In mathematics, we heard about how the solution of Fermat's Last Theorem or Cohen's Forcing came from such determinations. Obviously, "sheer" determination can not lead to real results, but I say that even deep determination can't lead to truly deep results. No offence to these mathematicians. In fact, I traveled half a world and waited half a year, working in Stanford's math library just to meet Paul Cohen. The fact remains that the mystery of the Continuum Hypothesis is still unveiled. Some believe that this new non heavens induced knowledge is all we can obtain. Materialism spread through the world, creating its phony idealism. The situation is bleak.

Waiting for the heavens to give new signs can test our true idealism. Amazingly, the old signs are always here. If we don't allow the new shallow Formalism to cover up history, then we have a lot to chew on.

Once a puzzle becomes personal, we get our calling. That's when the "drop everything" should set in. Yet it doesn't always happen. Best example, is Beltrami, who clearly saw the Logical consequences of non Euclidean Models and yet did not drop everything and dig toward Logic. On the other hand, Cantor is the perfect example of how one should behave.

And this brings me to the much more complex second problem, "the wisdom of idealism". If we see the light, if we feel the puzzle, we have to drop everything. This usually narrows our vision about the full universe. As long as we dig, this is not a problem, but most often the blindness remains. Most importantly, the success or unsuccessfulness of our digging is not important in this remaining blindness.

In a deepest sense, there is no success in the digging and breaking down the puzzles. Turning the question marks into exclamation marks, is not really a success. This deepest sense of Idealism, has been almost forgotten since Plato.

First of all, the meaning of "success" is controversial. A crazy person can justify everything in his own mind. But quite on the contrary, the social success can be an even more delusional confirmation. That's where the root of Social Formalism lurks. This is a land of total contradictions. The psychiatrists are the softest, but most subtle, tyrants of social insanity, yet the recognition of this, doesn't make L Ron Hubbard right. To stay more in scientific lines, the academic acceptance has nothing to do with truth. Just as in the social world, money can create "success", in the academic world, Formalism creates acceptance. All these are just matter of "times", unreal times, spreading of lies.

If the honest "digger" is not aware of these, then he'll become a sociopath. So after all, the falseness of society, creates a true falseness in the individum. It's a defense of society. It doesn't want truth to spread. The "wisdom" of idealism is the defense of the individum. It can be approached negatively as the collapse of "maya", the illusionary world of society. This is what usually is associated to yoga or enlightenment. An opposite approach is devotion to God, the most abused path today. The truest middle road is philosophy. True philosophy of course, is the truest yoga and truest devotion to God. But all these are abstract descriptions. There are only two elementary acts, namely to receive truth and to give truth.

The digger who sees a puzzle or the poet who hears the perfect lines, are receiving truth from the other side. He or she may not even believe in an other side though. To dig or to perform are secondary reactions. The primary should be to be overwhelmed by the existence of the other side. Without this, the secondary reaction already has a bad omen. And yet, truth spreads even in this distorted and mutilated form. But the messages are interpreted falsely. People become idolized and idols become poisoned.

The digging is aiming to know, to resolve the puzzle. Once it is resolved, it fits into the full world, which now becomes understood better. The explanation of all this to someone else must start from the full world, before and after the puzzle. This naturally leads to a falsification of what happened.

Most importantly, the solution induces new puzzles and this falsely can give the impression that nothing has been solved. This is not so, simply the first law in the "wisdom of idealism" applies, namely: The more we know, the less we understand.

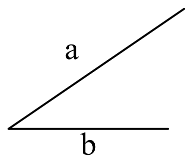
The discovery of "natural selection" by Darwin was the best example.

Contrary to widespread misrepresentations, Darwin was not the father of evolution, rather of natural selection. He knew this very well as the title of his book expresses: The Origin of Species by Means of Natural Selection. The idea of evolution was neither his nor a big discovery, it was expected by all sane people, but it represented a big puzzle. Namely, how could the complicated appearance and behavior of animals evolve through the narrow process of recreation. Nobody knew at that time about DNA, but everybody knew that sperms impregnate the eggs and embryos are formed. How can the sperms and eggs bring about the perfect species to their environment? The "natural" but false idea could be that the behavior of an animal is somehow affecting the sperms and eggs. For example, the giraffe who always reaches higher and higher, would produce sperms and eggs reflecting this tendency. To achieve this, through the blood stream, would be a total mystery, Darwin realized that what people did to breed animals offered a purely mathematical explanation instead. Just as the breeders select certain puppies to create new breeds of dogs, nature also selects certain individuals who live longer or mate more frequently. So the long neckedness of the giraffe didn't happen as a mystical reaction of reaching high to the leaves. Instead, the sperms and eggs have a correspondence to the way an individum looks or behaves. If a giraffe is five centimeters longer, then this difference has a corresponding difference in his sperms or her eggs. Since an individum including his sperm or egg develops from a single cell, this correspondence is not mystical at all. Of course, the development of an individum and its organs in general, is still a mystery beyond any concept of evolution. Accepting that the long necked giraffes have their long neckedness in their sperms and eggs too, creates new problems. If the sperms and eggs determine the new individuals, then they could cancel each other. The sperm might be long necked, but the egg not, or vice versa. Or they always attract each other at mating? Most importantly though, even if nature selects the long necked giraffes to live longer and mate, it would simply change the population to one that averages around the longest necked individuals. It couldn't grow from population to population. So it would fail to explain the whole original purpose, that is evolution. Indeed, we suspect that giraffes as a species evolved from an earlier, shorter animal. The solution is the concept of "mutation" This means that a combination of sperm and egg is not exact copy or even a combined copy. Two short necked giraffes can bring about a long necked one, though very rarely. Most likely, is that long necked parents create long necked children. So then at first, it seems that mutation is working against natural selection. But surprisingly, it is the true motor of it. It's true that it slows down a little bit the spreading of long neckedness in a population, but at the same time, it makes it possible that longest necked individuals can have offspring who will be even longer. Indeed, the longest necked sperms and eggs have mutations too, above or under. Now if the above is selected by nature more often than the under, then very slowly, the top range of the neck length, not only spreads, but actually increases from generations to generations. Besides this mathematical problem of how the mutation can become a motor, rather than a destroyer of the selections, a more fundamental problem is what features can evolve in this gradual manner. For example, the sharpness of a tiger's tooth is very similar to the giraffe's neck. But the bird, that opens eggs by throwing a big rock to it, couldn't evolve this behavior by throwing bigger and bigger stones. It is only perfect as it is. Not gradual. Similar is the nest building. Gathering twigs might be an ancient activity, but perfect circular nests can not be approached from useless forms. And then of course, the more complex behaviors are not even mentioned. To generalize the idea of natural selection to all life forms and all actions, requires to solve more and bigger puzzles than the original puzzle of avoiding transference to the sperm or eggs was. So, to deny natural selection is stupid. But to say that everything is caused by natural selection is even worse. It's suppressive stupidity. I went that far because it's exactly how it happens in math too. Solving one mystery, opens up a thousand more. This also helps to be more truthful in judging people, who went against the new discoveries.

When I fell in love with Set Theory and heard about the fact that some contemporaries of Cantor rejected him, I simply felt hatred against these "narrow minded" conservatives. Of course, the unfortunate fate of Cantor ending up institutionalized, helped me in this sentiment. And yet, now after seeing all the dead ends where Set Theory lead, I can contemplate the idea that those "enemies" of Cantor might have had a deeper foresight, than mere stupidity or ignorance. But we still have to stay on the right side, so lets start with Cantor's genius.

1. Irrationality and its generalizations

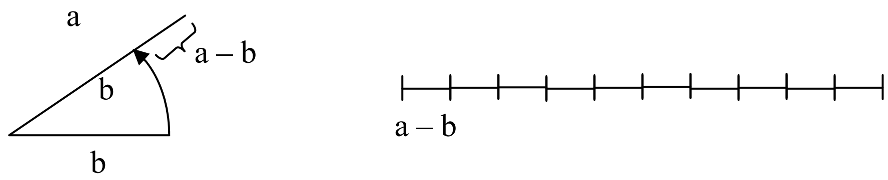
Numbers are merely the proportions of distances, or to put it another way, the measuring of one distance in units of another.



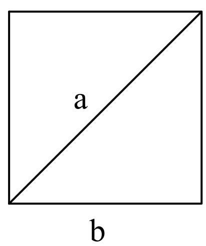
$\frac{a}{b}$ = a measured in b , being a unit

This ratio or measure of a by b can be expressed in more number like manner through our decimal system. We simply measure b into a as many times as possible. Like in the above picture, b fits only once into a, so $\frac{a}{b} = 1 \, . \, . \, . \, .$

To continue, we regard the leftover part from a and multiply it by ten:

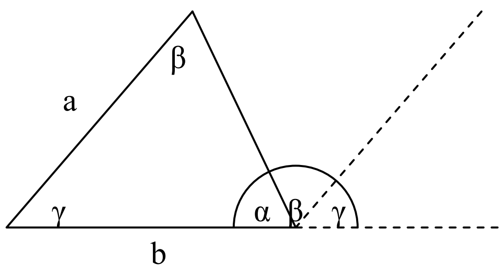


Then, we measure b into this as many times as possible. This can't be more than nine. Indeed, if it were, then a – b could have contained already b.
We always regard the leftovers and increase them ten times, to measure b into it again and again. For example, if the b base or unit is used to draw a square and a is the diameter, then:

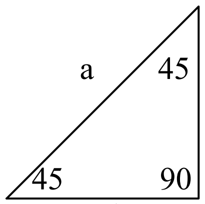


$\frac{a}{b} = 1 \, . \, 4142 \; . \; . \; .$

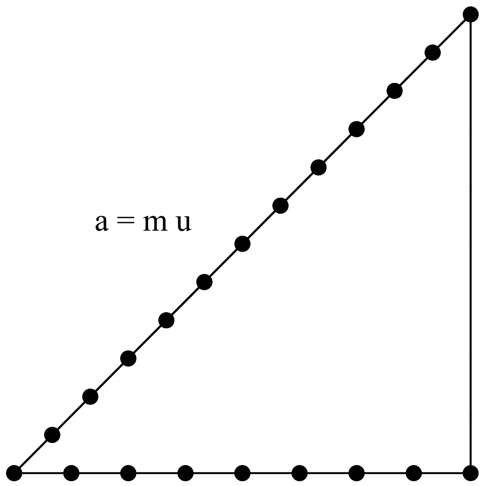
The fact that these distant ratios or numbers in units depend only on the shapes, but not the actual sizes is due to the proportionality of the Euclidean geometry. Indeed, blowing up pictures to larger size, the angles remain the same. Thus, after the points, lines and circles, angles are the real special concepts of geometry. The third, even more subtle level is parallelity. This combines all the complications. The parallels also shows that triangles are the most important and their angles in total are always 180^o :



So even our previous square diagonal problem is actually a triangle problem with 90 , 45 , 45 angles:



The decimal system is a silver platter, that makes this old problem easily servable. The Greeks themselves approached it differently. They asked whether we could measure a small u unit into both a and b exactly.



If that was the case, then $\frac{a}{b} = \frac{m\,u}{n\,u} = \frac{m}{n}$ so the ratio were a fraction.

It's no surprise that this is what they aimed for, because they didn't have the silver platter of the infinite decimals, that can compare any distances for sure.

The Greeks also realized that such common u unit does not always exist, namely it can't exist above either. So our picture above is a fake.

A second silver platter helps us today to see why not all ratios could be fractions.

Indeed, we all learnt in school how to divide fractions into decimal form, for example:

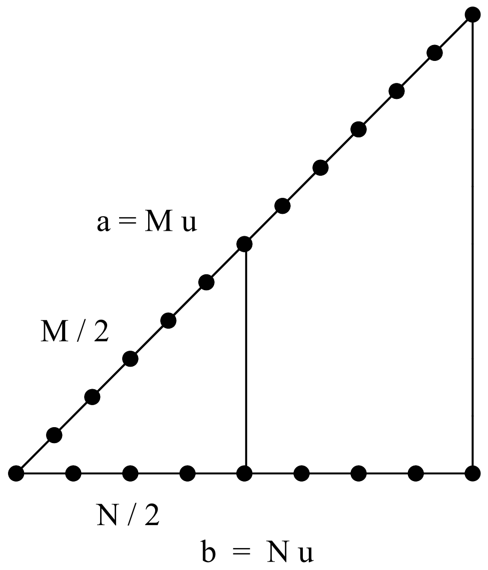
$$\frac{9}{7} = 9 : 7 = 1.\overline{285714}\,\overline{285}\,...$$

```
     2 0
       6 0
         4 0
           5 0
             1 0
               3 0
                 2 0
```

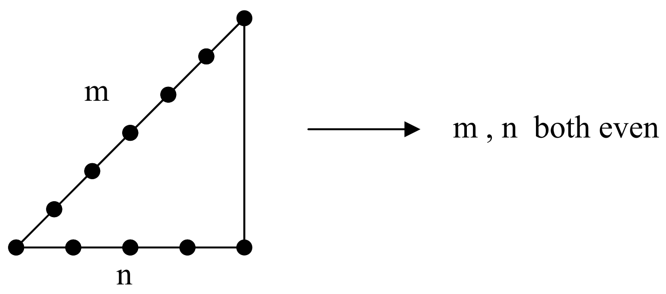
We simply applied the rule of decimality, that is the leftovers multiplied by ten.

This was achieved by putting zero after the remainders. Now the obvious fact is that if we divide with 7 we can only have remainders 1 , 2 , 3 , 4 , 5 , 6 , so after maximum

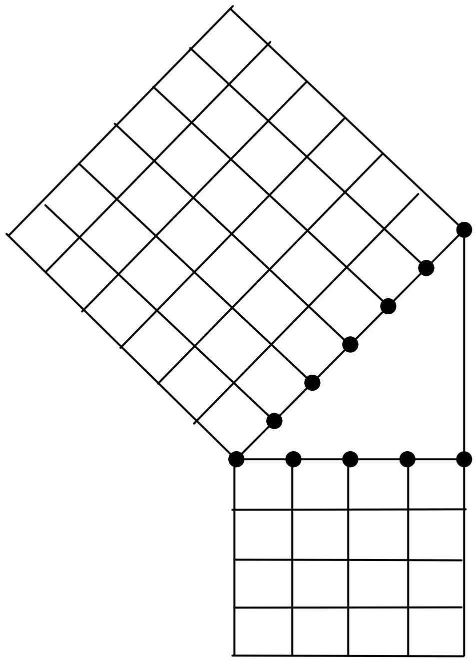
six divisions, a remainder must re-occur. Then everything repeats, so we always must get a periodic decimal form, but necessarily like above, repeating from right after the decimal point.
Arbitrary distances of course can bring about arbitrary infinite decimals. In fact, we can reconstruct the distances from the decimal forms by combining them from infinite many smaller and smaller pieces.
Since fractions only give periodic decimals, thus it's obvious that they only give very special distant ratios. So we shouldn't be surprised if the above $90, 45, 45$ triangle is such. An other way of saying that the $\frac{a}{b}$ ratio is not fractional is that it is irrational.
So our two silver platters, the decimal system and the division process provides us with the wisdom that irrationals are necessary. Yet it still doesn't show why the particular $90, 45, 45$ case is such.
To see it, first observe that if some $\frac{M}{N}$ fraction were the ratio of $\frac{a}{b}$ so that both M, N are even, then exactly by the proportionality of the Euclidean geometry, the same ratio were given by their halves $\frac{M/2}{N/2}$:



This could happen only because both M, N were even, so we could halve the distances exactly at units.
Of course, if $M/2$ and $N/2$ are both even again, then we can halve them again.
And so on, clearly, we must get a small triangle that will have m, n many units, so that they can't be halved again. In other words, they can't be both even. One must be odd. Now we show that this is impossible. That is, for any assumed m, n units they both must be even:



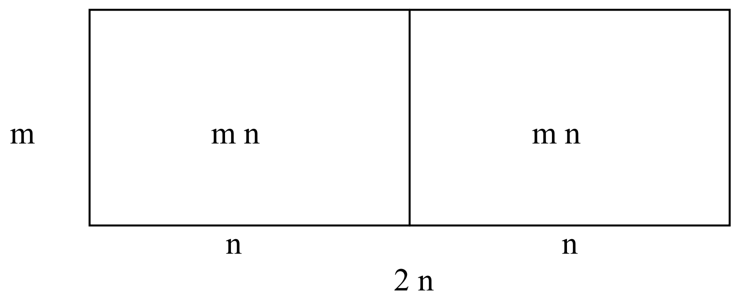
Lets put squares on both sides and tile them with unit squares.



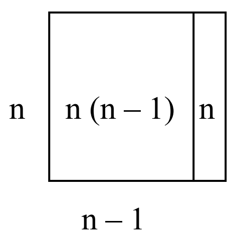
The number of tiles are of course m m = m^2 and n n = n^2.
The first general rule is that the evenness or oddness of the squares follows the sides.
In other words, $even^2$ = even , odd^2 = odd. We can confirm this by examples:
$2^2 = 4$, $4^2 = 16$, $6^2 = 36$, . . . while: $3^2 = 9$, $5^2 = 25$, $7^2 = 49$, . . .
A fairly obvious fact is that any m multiplied by an even, that is 2 n becomes even:

$$m\,2\,n = 2\,(m\,n)$$

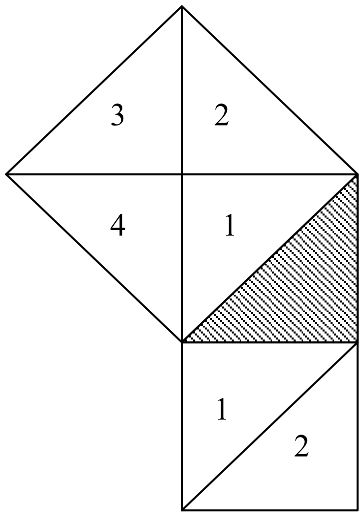
This can be seen with tiles too:



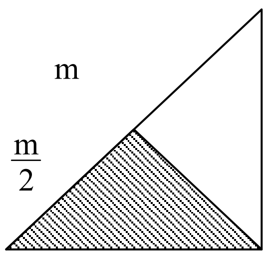
Thus, of course, an even even = $even^2$ = even is obvious too.
The inheritance of the oddness to the squares is a bit more tricky.
Observe that if n is odd, then n – 1 is even, so:
$n^2 = n\,n = n\,[\,(n-1)+1\,] = n\,(n-1) + n$ = even + odd = odd.
This can again be seen with tiles:



Now after the general law of inheritance, lets observe the special relation of our two squares, namely we claim that the big is double of the small:



As we see, the big square is four times the triangle, while the small is twice.
Then, with tilings we must have the same, that is $m^2 = 2\ n^2$.
Thus, at once we see, that m^2 is even, which by our general rule means that m is even too. Now to show that n has to be even too, lets half the triangle:



Since m was even, this keeps the units, so we got the smaller, similar 90 , 45 , 45 triangle with full units. But in this one, the n is the big side, so it must be even by our previous argument: $n^2 = 2\left(\frac{m}{2}\right)^2$.

This can be seen algebraically too from $m^2 = 2\ n^2$, if we write $2\ \frac{m}{2}$ in place of m:

$$m^2 = \left(2\ \frac{m}{2}\right)^2 = 4\left(\frac{m}{2}\right)^2 = 2\ n^2 \rightarrow n^2 = 2\left(\frac{m}{2}\right)^2$$

Thus, both m , n have to be even, contradicting that one was odd. This of course contradicts the original general M , N whole numbers as assumption.
If $b = 1$ is used, then $\frac{a}{b} = \frac{a}{1} = a =$ irrational. And of course, $a^2 = 2\ b^2 = 2$.
So, $a = \sqrt{2} = 1.4142\ldots$ is irrational.
By the time of Cantor, all this was obvious, in fact, the $a = \sqrt{2}$ form of a was generalized. Indeed, already a fraction, $\frac{m}{n}$ can be regarded as an equation:
$x = \frac{m}{n}$ is the same as $n\,x - m = 0$. Then, $x = \sqrt{2}$ means $x^2 - 2 = 0$.
We can regard higher order equations, like $x^3 - 3\ x^2 + 5\,x + 5 = 0$.

The solutions can obviously be irrational numbers, but can they be all possible numbers? The feeling was that even if we regard all possible equations with whole numbers, there might be ratios, that is infinite decimals that can't become roots of such equations. Then came Cantor with a new silver platter. Just as our silver platters of decimals and division process show at once that there have to be plenty of irrationals, Cantor's silver platter shows why there have to be plenty of decimals, that are not roots of whole equations. We would think that he was praised at once. The fact that his silver platter didn't go into the education system, shows that some things went astray. Just as an exercise, first we'll use Cantor's argument, not for its original application, that is for the roots of whole equations, rather merely to the fractions. We'll still use the silver platter of infinite decimals, but not the division process which shows that fractions are periodic. In fact, lets restrict even the decimals, just to the ones between 0 and 1, that is starting with a decimal point.
Lets entertain the idea whether we could list all possible such decimals under each other as:

.3 9 0 5 7 3 1
.2 1 0 9 3 7 2
.5 0 9 1 3 5 2
.
.
.

To claim that all possible decimals are listed above is pretty strange, but doesn't seem contradictory in itself. Now lets mark the diagonal digits:

.<u>3</u> 9 0 5 7 3 1
.2 <u>1</u> 0 9 3 7 2
.5 0 <u>9</u> 1 3 5 2
.
.
.

We can create a single decimal from these:

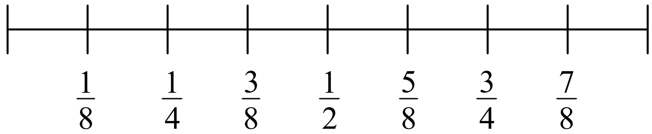
.3 1 9 . . .

Still, nothing weird shows, in fact, this diagonal decimal could easily be in our list.
But now, lets change every diagonal digit, for example, by adding 1 to it and for 9, meaning 0. Then, the anti diagonal number formed is:

. 4 2 0 . . .

This can't be in our list. Indeed, it can't be the first because the first digit of that, that is 3 was changed to 4. It can't be the second, because the second digit in that, that is 1, was changed to 2. And so on, it can't be the third, or any other.
So we created a definite missing number from the list. In other words, the listing of the decimals lead to the very impossibility of such full lists. This missing anti diagonal decimal was not a reflection on how many could be missing at all. It was merely the logic that collapsed on the assumption that there can be a full list.

So where does this lead us in showing that not all decimals are fractions? Well, we showed that the decimals are not listable, so if we could list all fractions between 0 and 1 , then it would be obvious, that some decimals are missing from such list.
The fractions between 0 and 1 are the real fractions, that is having smaller numerator and denominator. But still, the variations are plenty. In fact, they are dense on the 0 , 1 interval.



As we see, even just the halvings, that is the 2 powered denominator fractions, will densely spread in 0 , 1. To list all fractions one by one is still possible, namely by simply going in increasing denominators:

2 : $\frac{1}{2}$

3 : $\frac{1}{3}, \frac{2}{3}$

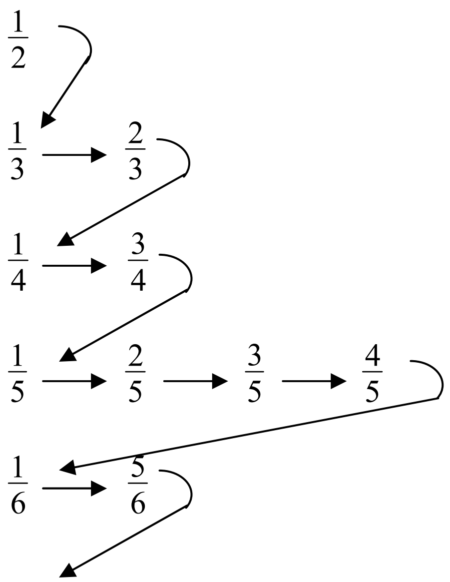
4: $\frac{1}{4}, \cancel{\frac{2}{4}}, \frac{3}{4}$

5: $\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$

6: $\frac{1}{6}, \cancel{\frac{2}{6}}, \cancel{\frac{3}{6}}, \cancel{\frac{4}{6}}, \frac{5}{6}$

.
.
.

I crossed out those fractions that can be simplified and thus, appeared already earlier.
We have a list, but not of individual fractions, rather, groups of them. Of course, each group only has finite many in them, so we can go through all of them as a single list, or if one prefers it, as a relisting.



$\frac{1}{2}$

$\frac{1}{3}$

$\frac{2}{3}$

$\frac{1}{4}$

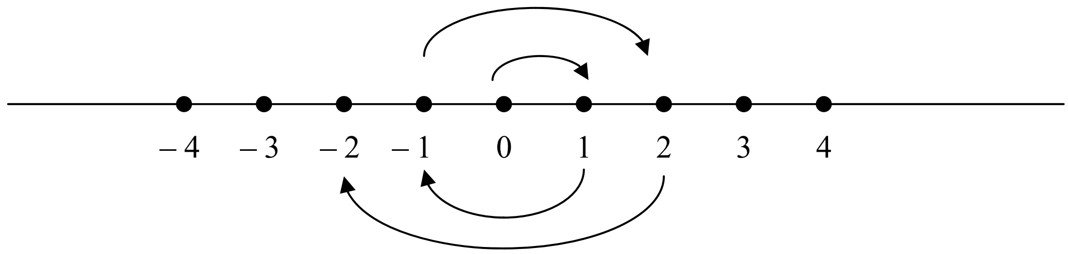
$\frac{3}{4}$

.

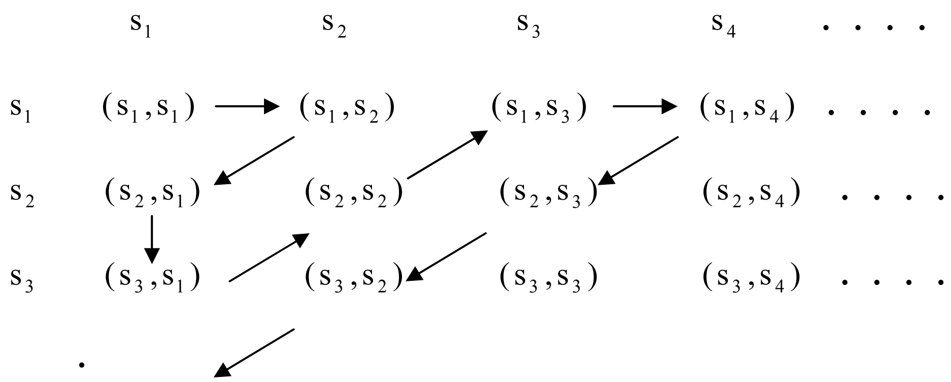
Each of the two proofs of Cantor, the non sequencability of the decimals and the sequencability of the fractions can make us dizzy. The two together as a proof for the existence of irrationals should definitely make us dizzy. Only going through repeatedly, shall we get a sense of reality.
But we have to go on, to the real argument Cantor used to prove with the same ease that there have to be decimals that are not roots of any whole equation.
An n-th order whole equation looks like this:

$$x^n + w_1 x^{n-1} + w_2 x^{n-2} + \ldots + w_{n-1} x + w_n = 0$$

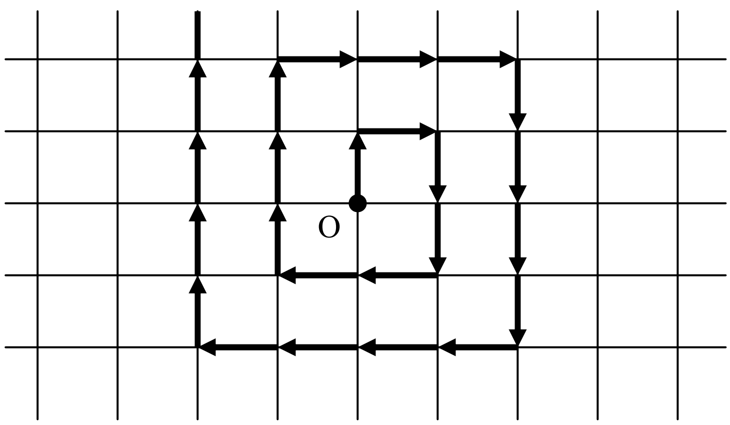
So it is determined by the $(w_1, w_2, \ldots, w_n)$ group of whole numbers. This is a kind of generalization of a mere fraction $\frac{m}{n}$ actually in two senses. Firstly, we can have negative numbers and we have more than two. The whole numbers themselves are easily a single sequence in spite of their two directionality:



Thus, the pairs of whole numbers are pairs of two sequences:



The arrows show how such pairs from two sequences can be made into a single.
Just as an interesting side "show", we can mention that the pairs of whole numbers can also be visualized as the grid points in a coordinate system. Then a spiral walk from the origin can sequence them all:



Now the argument for triplets of wholes (w_1, w_2, w_3) can go as follows:
A triplet can be viewed as a pair of one and a double:
$(w_1, w_2, w_3) = (w_1, (w_2, w_3))$
So since we already listed the single wholes and the pairs too, thus we actually have only pairs of two sequences. Similarly, we can list four, five, n groups of wholes:

List of doubles (w_1, w_2) : $(1, 1)$, $(1, 2)$, $(-1, -2)$, $(0, 2)$, . . .
List of triples (w_1, w_2, w_3) : $(1, 1, 1)$, $(1, 2, 3)$, $(0, 1, 1)$, . . .
.
.

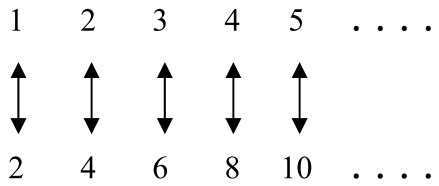
Thus again, we obtained a list of sequences, so with the same argument, even the total of these, that is all possible finite groups of wholes, can be put in a single sequence. Don't worry if you get dizzy, just repeat the arguments. They seem to be crazy and out of control, but they are not.
What we have shown is that all possible whole equations are merely a list. This still, has nothing to do with their roots. But it is well known in algebra that an n-th order equation can have maximum n roots. Thus, each whole equation has a finite group of roots:

Equations:	E_1	E_2	E_3	. . .
Roots:	$(r_1^1 \ldots r_{n_1}^1)$	$(r_1^2 \ldots r_{n_2}^2)$	$(r_1^3 \ldots r_{n_3}^3)$	. . .

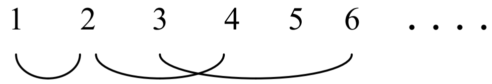
E_1 was n_1 order, so had only that many roots. And so on.
Then of course, we can go element by element, for each group, so we actually listed all possible roots as well. Thus, the decimals that are roots of whole equations are merely a single list, a sequence. So, they can't be the full set of decimals, because those can't be listed. In short, some decimals are not roots of whole equations.
A strange, sweet attraction and a bitter resentment is provoked from anybody by these arguments. The fact that some things so external to equations, roots and numbers, could prove facts about these is shocking to our intuitions. And yet, our intuition survive, in fact, can expand in this new universe of sets.

2. The Abstraction of Sets

As I mentioned, Cantor "dropped everything" when he encountered the power of sets. The sequencability or non sequencability was more than a new way to prove the existence of numbers that are not roots of whole equations. It was an entrance to a wider world. Pure sets lie under the colorful and confusing world of mathematical realities. The sequencability itself is also just a special case of something more general, "equivalence". Equivalence is when we pair two sets element by element.
Many people before Cantor realized the simple fact that comparing huge number of individuals like boys and girls in a ballroom can be done without counting. Indeed, all we have to ask is to form pairs. Whichever is left without pairs is more. Also, many people like Galileo realized that infinite sets are strange, because they can be equivalent to even only their own "half":



Indeed, we paired all numbers to their double. An other way to look at this is:



This is more confusing, even though, it shows more the actual inner pairings.
But it was only after Cantor's proof of the unsequencability of all decimals that this new infinite equivalence became useful. It showed that infinites are not a crazy world where everything is possible. There are yes and no-s too. If reality is a confusing, colorful disguise of sets, then the disguise of equivalence is isomorphism.
The isomorphism principle is an instant a-priori human ability to use a ruled system from one reality, to a new one. Every child who knows the rules of chess, will at once know the fact that playing with new different pieces can not affect the game. This principle is not elementary in the sense that it only "jumps in" after other intellectual steps had been activated. But it is elementary in the sense that it jumps in without reinforcement. Of course, it is also one of the best tests of human intelligence. Crazy "scientists" can fool even crazier amateurs, that animals can think, but they will never pass the isomorphism criteria. To claim that all structures are merely sets, and all isomoprhisms are merely equivalences, will probably turn out to be false. But at the present, we can't pinpoint why not. Cantor was definitely obsessed with the problem of equivalence. As time passed, the concept of sets became accepted, I even would call it as a silver platter today. But the concept of equivalence did not go along. In a sense, this is natural. The real mystery is why didn't the set concept, on its own, rise long before. Just think of Euclid's lines and circles, they cross in points. So what is more natural than to imagine the lines and circles as sets of points. When they cross, they simply share common element or elements. Of course, the greeks struggled to visualize how the points "are" on a line. But still, the points as elements could have risen. Once this formal set convention was accepted, the word set as collection became so natural, that today it became like a new geometrical counter part of the old trivial formal logic. Every lecturer now draws circles and says: "Imagine these are the people who this and this and this." This abstraction is annoying on one side, and

puzzling on the other, namely why it came so late. I don't see particular harm in how text books today say "the set of fractions" and so on. They use nothing about sets. They might as well just say "the fractions". By the aboves, it would seem that the problem is simply the forgotten equivalences. That those didn't penetrate the consciousness of the education system. But that's not the full problem. In fact, the real abstraction of sets was only crystallized after Cantor's obsession with equivalences became history.

Sets are mere collections of objects without any structure. A motorcycle as a set would be all the parts thrown into a basket, without any order. The more important positive aspect of sets is that we can put smaller baskets into bigger ones. So then the motorcycle can be organized without being put together. Just combining in one basket the directly relating parts, then in bigger baskets, the smaller collections and so on. Then, it's not so surprising that while physics is about how the parts interact, math is only about how they relate to each other.

When we start to formalize sets in language, then this seemingly natural collection vision gets harder, because we write in a line. So to say, that the set containing the numbers 1 and 2 is $\{1 , 2\}$ is misleading, because we wrote them in this order. The only thing we can do is to say that $\{1 , 2\} = \{2 , 1\}$. So we go back from the ordered language to the non ordered collection by claiming equality.

What is more basic, language or sets?

Logic proves that thinking is not happening in the natural languages. Grammatics overrules grammar. See article Grammatics. But sets are still a big problem, so as I said earlier, the critics of Set Theory were right in something. In spite of all that, there is no more fundamental symbol in the universe than $\in$, the relationship of belonging or being element:

$1 \in \{1 , 2\}$, $2 \in \{1 , 2\}$, $1 \in \{1 , \{1 , 2\}\}$, $1 \in \{1 , 2\} \in \{1 , \{1 , 2\}\}$

As we see, language allows to use the same object repeatedly, and also sets can use the same object at different levels of collection.

But in one set, we can't repeat elements! We can repeat it in the language of collection, but they collapse in the actual sets: $\{1 , 1\} = \{1\}$

This also shows that an "abuse" of sets must be applied to reflect language. In other words, if we want to allow the collection of repeats or order, then we have to do it in different levels, merely as a trick. So for example, $\{ \{ \ \} , \ \ \}$ could be a trick to collect pairs that have an order and can even repeat:

$\{\{1\} , 2\} \neq \{\{2\} , 1\}$ $\{\{1\} , 1\} \neq \{1\} \neq \{\{1\}\}$

All these show also that from one single object, we could build a whole universe.

So it's not the what, rather the how that counts.

The usually accepted way to define the (x , y) ordered pair is $\{\{x , y\} , x\} = (x , y)$. So the first number is "repeated" A set of ordered pairs is the R relation.

$(x , y) \in R$ is also abbreviated as $R (x , y)$, to reflect the original meaning.

This expresses the fundamental superiority of sets. We are not interested in what creates the relation between x and y, we merely establish the fact of such relation by collecting all (x , y) pairs that are in relation. This is like sociology versus psychology. We don't deny the psychological factors, but examine the mere facts in a group of people.

If in a relation, for every x, there is only one y, so that $(x , y) \in R$ then such R is called an f function and instead of the $f (x , y)$ abbreviation, we prefer $y = f (x)$.

This then pushes the set meaning further, but allows the subjective meaning to be expressed too. Indeed, we feel that f is "producing" the y value for every x.

This subjective meaning of functions is even more emphasized by the concept of domain and range. This can be defined already for $R (x , y)$ as the collection of all

appearing x-s and all appearing y-s. But here, at functions, the domain, that is all appearing x-s are regarded as the set where f is "defined". The range is the set of values that f "picks up". These subjective meanings help, if used precisely.
Best example, is for the equivalence. Initially equivalence could be simply defined as an R (x , y) relation, where not only y is unique to x, but in reverse, to every x there is unique y. But in an other more specific way:
y = f (x) is an equivalence between the X and Y sets if:

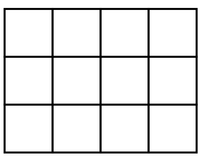
1.) For every $x \in X$, f is defined, that is y exists. (X is the domain)
2.) Not only y = f (x) is unique, but x is too.
   This means that different $x_1 \neq x_2$ give different $f(x_1) \neq f(x_2)$.
3.) For every $y \in Y$, there is x that y = f (x). (Y is the range)

Then to prove the crucial Cantorian second feature of X < Y, that is the impossibility of equivalence between X and Y can be done by showing that for an f that satisfies 1.) and 2.), the 3.) will fail. That is, an equivalence "on" X can't give the full Y. The range can't be Y. Of course, allowing a non equivalent function, that is same y values for different $x_1 \neq x_2$ or a function with smaller domain, will make it even less possible to obtain the full Y. So, it's enough to show that for any function defined in X, the range can't be Y. It's enough to refute 3.).
All these subjectivized set constructions raise the problem of whether we use sets to merely exactify natural claims or to prove surprising results. Of course, with Cantor's results, we see that it's a mixture of both. This situation is very different from number theory or geometry and we'll come back to these questions in detail.
The most clear, "explicit" way of defining new relations between sets is to express these logically from the fundamental $\in$ relation. Simplest such is the "subsetness":
$x \subseteq y : \forall z (z \in x \rightarrow z \in y)$
$\forall$ means "for all" and $\rightarrow$ means "imply". So, x is subset of y, if all elements of x are elements of y. The line underneath $\subset$ shows that it's possible that x and y are equal, namely if in reverse $y \subseteq x$ too: $x \subseteq y \wedge y \subseteq x \rightarrow x = y$.
$\wedge$ means and and $\rightarrow$ implication again. $x \subseteq y \wedge y \subseteq x$ can be also expressed as $\forall z (z \in x \leftrightarrow z \in y)$. So what we say is that if x and y have the same elements, then they are the same. The fact that in the condition we don't use equality, only in the consequence, shows that something is fishy. And indeed, we have a fundamental problem. Should we define equality by having the same elements? Or is equality a wider concept? This involves Logic, but most importantly, boils down to the question whether we allow sets without elements at all. If so, then the equality or difference of these elementless sets or so called urelements can not be defined by the elements. For example, the 1 , 2 , 3 , . . . natural numbers seem to be urelements.
But, as we said, we can build up repetitions from a single element too. So, the practice became that the only elementless set is <u>the</u> empty set $\varnothing$ and everything else, including the naturals are built from this. Then, indeed, equality is reduced to having the same elements.
The $x \subset y$ or proper subset relation simply means that $x \subseteq y$ but some elements of y is not in x: $x \subset y : x \subseteq y \wedge \exists z (z \in y \wedge z \notin x)$
Here $\exists$ means, "there is such". Observe that $\wedge$ means and but also "but".
$z \notin x$ is a short for not being element, that is $\neg (z \in x)$.

3. The Explicit Collection

The most elegant way of using the explicit logically built relations from $\in$ is not to define new relations, rather to collect elements, that is form sets. We use the same special { } brackets for this as before, but we use z ; in the beginning. This z is the dummy variable for the collected elements, while the other variables determine the collection. Simplest is the formation of the combined set or union of two sets:
$x \cup y \;:\; \{z \,;\, z \in x \;\vee\; z \in y\}$ Here $\vee$ means "or".
The intersection or common part is the collection of the common elements:
$x \cap y \;:\; \{z \,;\, z \in x \;\wedge\; z \in y\}$
So it seemed that this $\{z \,;\, . . .\}$ is an other basic symbol of Set Theory besides $\in$.
In fact, it also seemed that this $\{z \,;\, . . .\}$ could be axiomatized by the simple rule:
$z' \in \{z \,;\, . . .\} \;\leftrightarrow\; . . . z = z' . . .$
Indeed, a z' should be element of the collection, if and only if, using z' in place of z, the . . . . . is true of the collection. For example, a $z' \in x \cup y$ if and only if, z' is an element of x or y, so indeed $z' \in x \;\vee\; z' \in y$.
Certain things are true for all objects! The simplest is $x = x$.
So then collecting all objects that are equal to themselves: $U = \{z \,;\, z = z\}$.
This should be the whole universe! But since everything is equal to itself, thus, the universe is too, that is $U = U$. So, U itself obeys the collection property $z = z$.
Then by our axiom, U should be an element of the collection. $U \in \{z \,;\, z = z\} = U$.
So U is element of itself! Since U is the universe, we could forgive this absurdity from it. But more serious problem arises if we try to separate the normal, that is $z \notin z$ kind of sets. Their C collection can't neither be normal, nor abnormal:
Indeed, let $C = \{z \,;\, z \notin z\}$ then:
$C \notin C \;\rightarrow\; C \in \{z \,;\, z \notin z\} = C$ contradiction.
$C \in C \;\rightarrow\; C \notin \{z \,;\, z \notin z\} = C$ contradiction.
Since both of the $C \notin C$ and $C \in C$ options are contradictory, thus, the axiom of collections is itself contradictory, even if we would allow collections to be element of themselves. This example that lead to contradiction is the famous Russell paradox. It gave a death blow to the collection axiom and for a "second", to Set Theory itself.
The solution became to abandon the use of $\{z \,;\, . . .\}$ for arbitrary . . . and instead specify special splintered collections as axioms. In practice, however, the usage of $\{z \,;\, . . .\}$ remained with the promise of not causing contradiction.
The power of set collections is best shown by how we can generalize the union and intersection to set of sets. In other words, we regard the elements of an S set as sets, not just elements, that is we consider their elements too:
$\cup S$ is the union of all S elements as sets: $\cup S = \{z \,;\, \exists s \,,\, s \in S \wedge z \in s\}$.
$\cap S$ is the intersection of all S elements: $\cap S = \{z \,;\, \forall s \,,\, s \in S \rightarrow z \in s\}$.
The union is sometimes also called as sum and $\cup S$ is denoted as ΣS. This suggests that a more "algebraic" pair of it exists too, namely the product set ΠS.
Multiplication is the repetition of addition: $3 \bullet 4 = 4 + 4 + 4$.
But there is a more direct way too, namely as the area:

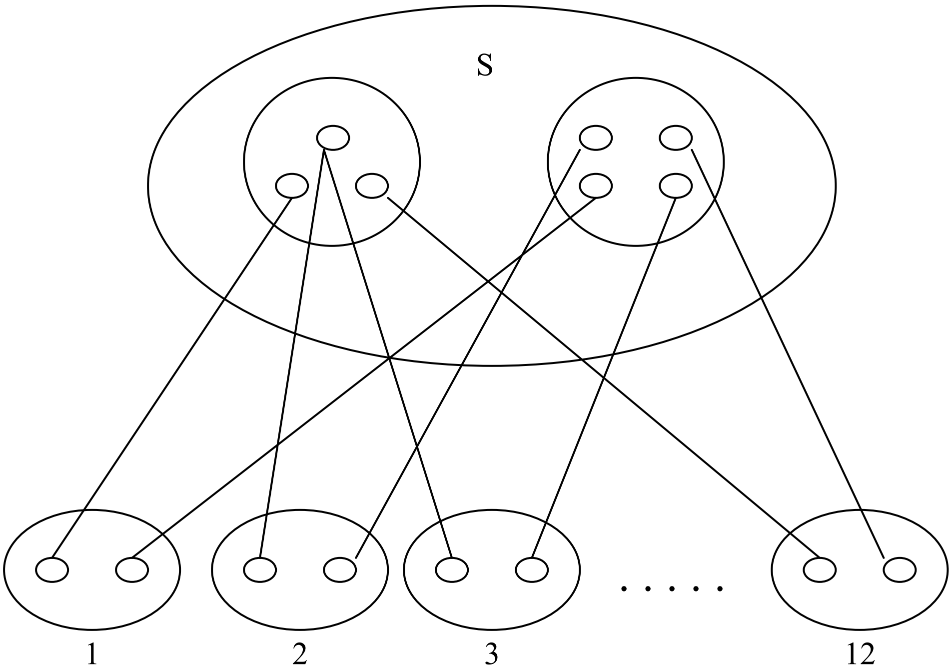
3



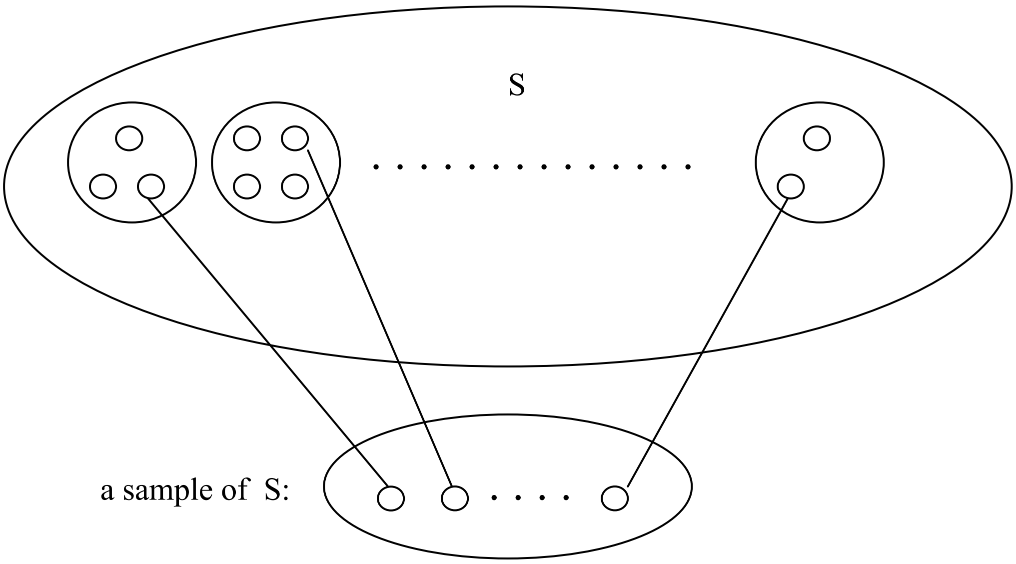
4

Every little tile is made from a pair of one element of the 3 and one of the 4 sides.

For sets, we merely regard one that has two elements, one with 3 and one with 4 elements. Then picking the pairs, give exactly the product of the two elements:



We went in random order, but it's obvious that there can be twelve pickings.
For an S set, with arbitrary many elements, then similarly, a sample of S is merely a collection of elements, so that one is picked from every element.



There is a big problem though! As we said, same sets can be collected in different sets. Thus, S might have s , t elements, that have a common c element. Picking c from s while a d different one from t will contradict the vision of the sample, because two elements will be collected from t, namely c and d. Also, if we pick the same c from t, then not every pick will represent an element of S, namely c represents both s and t.
If we assume that all elements of S have no common elements with each other, that is they are "disjoint", then the sample forming is indeed arbitrary picking.

But there is an other way with which arbitrary picking can be applied, even for non disjoint elements. The idea is to tell not just what we pick but also from which element of S we picked it. So then the sample won't be a set, rather an f function, that gives for every s element of S, the picked f (s) element of s.
Then if two s , t elements of S have a c common element, and we pick this c from s, it will merely mean that f (s) = c. So picking even this same c from t, will only mean that f (s) = f (t). But the "picking" now is not c rather (s , c) and (t , c), and these are not equal. So, the number of samples in this functional sense, keeps the multiplication. Thus, the set of all such functional samples should be ΠS:
$\Pi S = \{z \; ; \; z = \text{sample function of } S\}$
Now comes the real surprise!
If all the S elements have at least two elements themselves, then the product set of S is a set that is definitely bigger than S. In short, $S < \Pi S$. So we found the generalization of Cantor's basic idea to create new, non equivalent infinites.
The fact that ΠS is at least as big as S, is quite obvious. Indeed, lets pick a sample randomly and then change it only in one of its pickings! Depending on which pick we change, we get as many new samples as many elements S has. In other words, these "almost" same samples to the randomly picked one, give as many variants as the elements of S. Here, "almost" meant altered in only one pick. This also shows that altering more picks simultaneously, should give much, much more. But this subjective plentiness of the possible choices or samples is not enough. We need a logical line that shows for every pairings of S elements to samples, that the pairings can not involve all samples. Some had to be left out. So, let f_s denote the sample paired to s. This f_s picks from every t element of S an element, namely $f_s(t)$. In particular f_s picks $f_s(s)$ from s. These are the "diagonal" elements in the elements of S. Since each element of S has at least two elements, we can alter $f_s(s)$ to an other $\overline{f_s(s)}$ element of s. Let these $\overline{f_s(s)}$ elements ordered to the S elements as pickings be the sample g. We claim that this g is missing from the f_s ones. Indeed, from every s, the g (s) is picking the $\overline{f_s(s)}$ different from what f_s picks. Thus, g (s) can't be any of the f_s.
This sample result to create the bigger ΠS set than S, gives the impression that it is depending very much on the elements of the elements of S rather than on S itself. Of course, we said that the only assumption was to have at least two elements.
So, we can create a new set from S by putting exactly two elements, say $\{0 , 1\}$ into them. Or if we wish, we can attach this same $\{0 , 1\}$ to every element of S.
These $\{0 , 1\}$ samples will be a bigger set than S. This still involves these elements, not merely S. But now, we can give a new meaning that will change all that. Indeed, for an f sample, f (s) = 0 can be interpreted as "no s", while f (s) = 1 as "yes s". This of course is meaningless, because an s element can't be no or yes, but a collection of s can be. Then, every f sample defines exactly a unique subset of S. Namely, we regard those s elements where f (s) = 1 = yes. So then what we proved is that the set of subsets in S is a bigger set than S itself. This argument also shows why 2^S is used as an abbreviation for the set of subsets in S. Then in short, $S < 2^S$.

4. The Inductive Collection

This idea brings us to something that I already mentioned earlier about the intuitiveness of claims. This angle is very fundamental, yet it's never emphasized on its own. The axioms of a system are chosen to express the assumptions that we regard as obvious. Then, we derive by Logic, consequences that are not obvious at all. These are the theorems. For Number Theory, this is a pretty fair description of the truth.
The addition and multiplication are defined by axioms that determine their values intuitively from the basic counting or consecutiveness relation. Then, we can define more complicated things, like primes and claim very surprising results. Of course, besides the defining axioms of the addition and multiplication, we need the heuristic axioms that express the following: For every $R(x_1, x_2, \ldots, x_n)$ logically built relation, if we fix all variables, except one say x_i then either there is no such x_i at all that R would stand, or if there are, then there is a smallest that we can denote as $\mu x_i R(\ )$. Indeed, among any collection of numbers, there has to be a smallest.
The reason we don't want to express the axiom this way, is that we only want to talk about numbers, but not sets. So, $P(x)$ logically built property could replace the sets of numbers. But we don't want to use concrete fixed numbers only in a $P(x)$, rather temporarily assumed fix values in an $R(x_1, x_2, \ldots, x_n)$. This gives the freedom of changing those values. So, $\mu x_i R(\ )$ is actually a function of $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ giving the minimal x_i as values.
A more disguised version of this minimality axiom is the so called induction axiom that claims the following. If $R(\ )$ stands for $x_i = 1$ and from every x_i value, the relation inherits to $x_i + 1$ then it stands for all values:

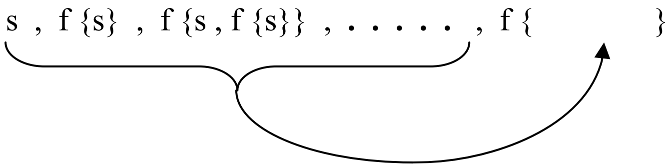
$$[R(x_1, .., 1, .., x_n) \wedge (R(x_1, .., x_i, .., x_n) \rightarrow R(x_1, .., x_i+1, .., x_n))]$$

$$\rightarrow \forall x_i R(x_1, .., x_i, .., x_n)$$

The name refers to the fact that $\forall x_i R$ was obtained from 1, step by step, that is "inductively", one truth inducing the next. The idea of induction is meaningful not just for truth, but for object collection too. In fact, it's very frequent amongst numbers. The famous Fibonacci numbers are always the sums of the last two, starting from 1 , 2. That is: 1 , 2 , 3 , 5 , 8 , 13 , 21 , . . .
This shows that the next element doesn't have to be determined from the last.
In the most general case, the new elements can be determined by the totality of all earlier ones. Among numbers, this is not emphasized, because as we said, we try to avoid sets anyway. But among sets, this idea gives a heuristic window to generalize inductive determination. Indeed, getting new elements from last, or even from last few, would merely give an infinite sequence. But using all achieved ones to produce a new, means that we can continue after a whole sequence. We merely have to have a way to order new element to this sequence as an infinite set too. Then of course, new infinites can be induced, step by step again. Usually, this side of the new idea is emphasized, that is obtaining the infinites gradually. But there is an other, much more important side. First, lets still remain with the old fashioned view and emphasize an elegance in the (set → new element) induction.
Among numbers, it's already possible that a new is the same as an old. Of course, if only the last determines the new, then this would mean a cycle:

$$1\ ,\ 2\ ,\ \underbrace{3\ ,\ 7\ ,\ 9\ ,\ 5}\ ,\ \underbrace{3\ ,\ 7\ ,\ 9\ ,\ 5}$$

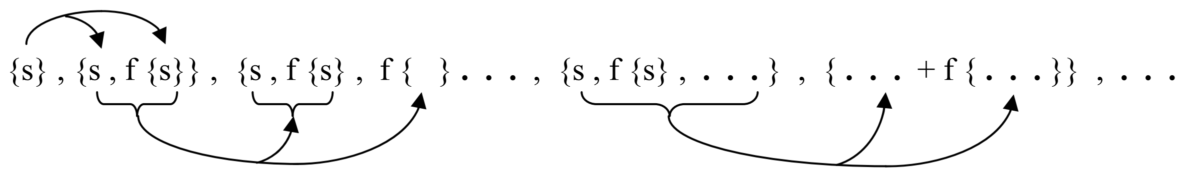
With more complex dependence on previous elements, the repetitions can be more complex too. If however, the new elements are ordered to the full set of the earlier ones, then we can only have one scenario. Namely, a new element that has already appeared. Then this single new would be repeated forever, because the total set remains the same. So in fact, we shouldn't even regard the new elements, rather the widening achieved ones. Then, the fairly stupid idea of infinite fix repeat can be omitted and instead a final full achieved set is regarded only. This means that we have a heuristic universal method for any f function, by simply adding the function value to the already achieved S sets or "stages". If S is a stage in our collection, then
$S + f(S) = S \cup \{f(S)\}$ is the next stage. This can only stop at S for two reasons:
Either f is not defined on S or $f(S) \in S$ so, $S + f(S) = S$.
But this grand view still needs two vital details.
First of all, we need an s starting element. Of course, f is not ordering the next element to s, rather to {s}, that is to the single set of s. Indeed, f gives the new numbers from the achieved sets and starting with s, the achieved set is {s}. Then:
s , f {s} , f {s , f {s}} , . . . , is the sequence of new elements, or in a better way:
{s} , {s , f {s}} , {s , f {s} , f {s , f {s}}} , . . . is the sequence of widening achieved sets or stages.
If f doesn't stop or repeats after finite many elements, then we get an infinite set.
If f is not defined on this or gives an element of the set, then this is the total stage.
But, if f gives a new value, then we can continue:



In the widening picture, the infinite set is definitely listed, regardless whether f can continue it or not.

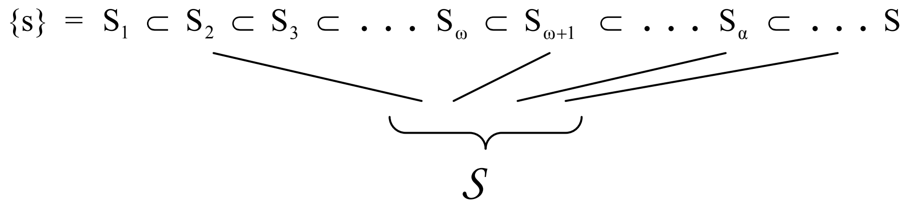
{s} , {s , f {s}} , {s , f {s} , f {s , f {s}}} , . . . , {s , f {s} , . . .}

This last stage was the total of the increasing earlier ones. If f does continue, then:



As we see, the widening "stage picture" is even better here. The new f value always comes from the last stage. In fact every new stage is the last stage, plus its f value, or if there was no last stage, then it is merely the combined earlier stages.
With the s start and the idea of combining, every f function gives a self determining widening that stops by f reaching an F stage where either f is not defined anymore or gives an f (F) value that was already picked, so adding it to F is useless.
The F final stage is totally determined by f and s.

The set of all stages will be denoted as $\mathcal{F}$ and here the letter is not referring to "final" as in F rather to "full". Neither F nor $\mathcal{F}$ are describing the f function. Indeed, f can be much more complex and the F final stage and $\mathcal{F}$ full set of stages, are specific to the s start. For different starts, we would get different stages.
$F \in \mathcal{F}$, in fact F is the widest element of $\mathcal{F}$.
If we regard an arbitrary $S \in \mathcal{F}$, then similarly we can regard all the stages that lead upto S. The set of these could be denoted as $\mathcal{F}(S)$. If S had no previous stage, then S was the combined set of these, that is $S = \cup\, \mathcal{F}(S)$. Most importantly:
Any $\mathcal{S}$ set of stages that "reaches up to" S, that is has no widest element smaller than S, will have the same total, that is, $S = \cup\, \mathcal{S}$:



But if $\mathcal{S}$ had a widest element, then of course, $\cup\, \mathcal{S}$ is this widest element.
So, we obtained a totally universal rule for any $\mathcal{S}$ subset of $\mathcal{F}$:
$\mathcal{S} \subseteq \mathcal{F} \rightarrow \cup\mathcal{S} \in \mathcal{F}$
This is the only property of $\mathcal{F}$ that is not depending on f or s.
The two properties depending on these are obvious. Namely {s} must be in $\mathcal{F}$ and f must give a stage from any S stage if f is defined on S.
So the three rules of $\mathcal{F}$ are:

1.) $\{s\} \in \mathcal{F}$.
2.) $[\, S \in \mathcal{F} \wedge$ f is defined on $S\,] \rightarrow S + f(S) \in \mathcal{F}$.
3.) $\mathcal{S} \subseteq \mathcal{F} \rightarrow \cup\mathcal{S} \in \mathcal{F}$.

So now, besides the visual conviction that f and s determine a full spreading or widening, we have these rules and so, we can say that $\mathcal{F}$ is the set obtainable by these rules.
Of course, these rules only tell what must be in $\mathcal{F}$.
They don't exclude any other rubbish we would add.
So now, we could even rise above these specific rules and ask in general:
If we give any rules for membership that specifies some elements directly, like above rule 1.), or conditionally, like 2.) , 3.) giving new elements from some old, then can we claim a result set, that is the collection of exactly these rules and nothing more?
It doesn't take much practice with sets to realize that the "rubbishless" collection from rules can in fact be reduced to explicit collections. Indeed, we can regard merely the

"completeness" as obeying some rules. For example, we could say that a z set is "stage complete" from f and s if:

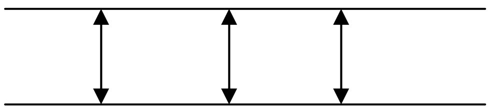
1.) $\{s\} \in z$.
2.) $[S \in z \wedge f$ is defined on $S] \rightarrow S + f(S) \in z$.
3.) $\mathcal{S} \subseteq z \rightarrow \cup\mathcal{S} \in z$.

Then, such z can contain rubbish. But now comes the trick. We can regard the common part of all these stage complete z-s to get $\mathcal{F}$. That is:

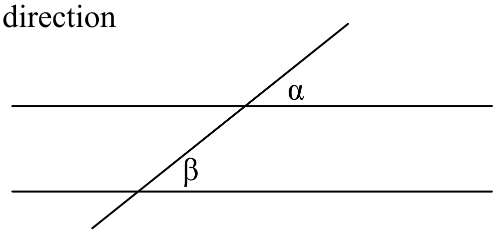
$\mathcal{F} = \cap\ \{z\ ;\ z$ is stage complete, that is: 1.) $\wedge$ 2.) $\wedge$ 3.) .

Indeed, the common part can only contain the rubbishless minimum.
Before we dwell into why this idea failed, lets go back to the heuristic wider problem of axiom systems. As I showed in the beginning of this section, in Number Theory the view that the axioms are the only obvious claims is quite convincing.
In geometry, it's definitely not the case.
The crucial concept of parallelity has three meanings:

1.) Same distance



2.) Same direction



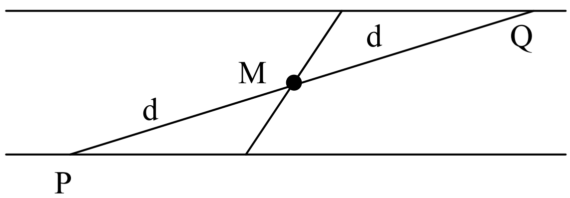
$\alpha = \beta$

3.) Non crossing

never cross

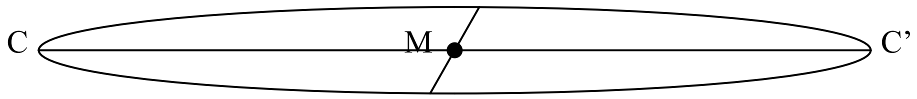
never cross

Obviously, 1.) implies 3.). Less obvious, but not hard to prove is that 2.) implies 3.) too. Indeed, if we mirror any point of one of the lines to the middle point of their connector, we get a point on the other line.



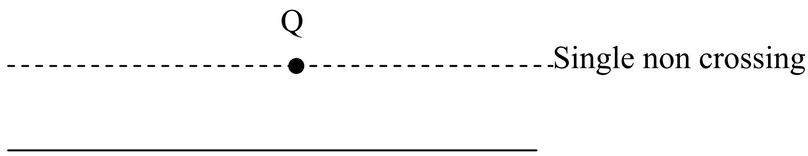
Now if the two lines had a P common point, then the Q mirror of P were an other common point. So then the two lines had two common points.

This is a contradiction. Indeed, two points determine only one line, but in our situation we would have at least three, namely the two parallels, and a mirroring line:



Unfortunately, there is a little bit of error in this proof. Because, what if P-s mirror image to M, is again P itself. Then, there is no two points, so then there could be two lines. Then of course, each line had the common point accessible, both to the left and to the right, from any starting point. But geometry has also axioms that ensure, that the two direction of a line from a point to left and right can't meet again.
It's still good to see, that if we drop this assumption of two directional infinity, then 2.) wouldn't imply 3.).
The real important question for Euclid was how to prove that 3.) implies 2.). That is: Non crossing lines are same directional to any connector.
In a negative form, this means that if the α , β directions are not equal, then there is a crossing. Euclid even specified, that for $\alpha > \beta$, the crossing is on the right side.
Since he couldn't prove this, he accepted it as an axiom as his infamous fifth postulate.
Of course, an axiom should be more simple than specific. If there were an axiom that is simple plus specific then it would be even better. If in addition it would only concern one of the three visual meanings then there could be no argument about its superiority. There is such form, as follows:

For any given line there is only one non crossing other, going through any outside Q point:



Indeed, any P point of the given line, connected with the outside Q can be regarded as a potential connector. The equidirectional lines going through Q are all non crossing. But, if there is only one such non crossing, then this single non crossing must be equidirectional. Otherwise the equidirectional couldn't be a non crossing.
This argument is a bit tricky because it assumes the possibility of more non crossings temporarily. Probably this is the reason that Euclid avoided this indirect line of reasoning, and only two thousand years later was it discovered. We don't even know exactly by whom, only that John Playfair wisely used it and claimed to learn it from others.
Euclid was not happy with his axiom and I'm sure he would have been happier with Playfair's. Strangely many mathematicians weren't happy even with Playfair's, so the search went on to try eliminating it, which lead to the non Euclidian Geometries. This in turn proved that some axiom for parallelity can not be eliminated. This was the intricate argument that actually showed that a Logic and Models as realities exist for math.
Returning to geometry itself then, it seems that the choice of an axiom for parallelity is arbitrary. The vision of parallelity is wider than any choice. In spite of this, as we go further and further in geometry, the theorems do become more and more surprising.
So it also seems that intuitively obvious theorems are only temporary "disturbances" in the choices of axioms.

A whole new situation arose after calculus became analysis. With the precision of limits and continuity, things that seemed natural before, like crossing and minimum or maximum of functions, became provable. At the same time, surprising, non intuitive truths were discovered too. With analysis becoming topology, this problem widened even further and yet never became addressed. Of course, Set Theory includes everything, so the ultimate question is, how or why complex Set Theoretical theorems can become intuitively simple. If we start at the bottom with explicit collections, then the Russell paradox was a bad omen. Yet, the inductive collection introduced above shows that intuitiveness reappears in complicated constructions. This is quite contrary to how in Number Theory intuitiveness only appears in the axioms. Strangely, the inductive collection was a generalization of the number induction.
The beautiful result above was that the intuitively obvious F widest stage in the widening of an f function from an s start can be obtained explicitly as:
$F = \cup\ (\cap\ \{z\ ;\ 1.) \wedge 2.) \wedge 3.)\ \}\ ) = \cup\ (\cap\ \{z\ ; z \text{ is stage complete }\} = \cup\ \mathcal{F}$.
In other words, the subjective vision of step by step, that is "timely" collection can be eliminated. Set Theory rules.
Unfortunately, we failed! Namely, we can't prove that there are stage complete z sets at all. Thus, $\mathcal{F}$ and so F too, could be the empty set. This is very surprising!
Indeed, we "feel" that the widening stages starting from {s} gradually "fill up" $\mathcal{F}$.
But since $\mathcal{F}$ claims to be the full set of all possible stages, thus, it can't be guaranteed to be anything at all. This suggests a solution too. We should regard the gradual widenings, so instead of reaching for the total $\mathcal{F}$ by, narrowing the hypothetical stage complete z sets, we could build $\mathcal{F}$ up from partial versions of it.
These are the $\mathcal{W}$ widenings. Almost behaving like $\mathcal{F}$, but the widest stage in them doesn't have to be continued by f and even a widest element doesn't have to exist through union. Thus, 1.) can be kept exactly, but 2.) and 3.) must be exempted for a stage, if it is the widest in $\mathcal{W}$ or were the total union: So $\mathcal{W}$ is a widening if:

1.) $\{s\} \in \mathcal{W}$.

2'.) $[\, S \in \mathcal{W} \wedge$ f is defined on $S \wedge S \neq \cup \mathcal{W}\,] \rightarrow S + f(S) \in \mathcal{W}$.

3'.) $[\, \mathcal{S} \subseteq \mathcal{W} \wedge \cup \mathcal{S} \neq \cup \mathcal{W}\,] \rightarrow \cup \mathcal{S} \in \mathcal{W}$.

4.) ?

Now what is this 4.) ? mystery condition? This is what we need to exclude the rubbish. Indeed, we can't use $\cap$ for that now. It would merely give the {{s}} shortest widening.
The most surprising fact is the following: If the $\mathcal{T}$ total of all $\mathcal{W}$ widenings satisfies 1.) , 2'.) , 3'.) , then it actually satisfies 1.) , 2.) , 3.) too.
Indeed, $\cup \mathcal{T}$ must be element of $\mathcal{T}$ otherwise adding it to $\mathcal{T}$ we would obtain a bigger widening. Also, f can't continue this $\cup \mathcal{T}$, that is either it's not defined on it, or $f(\cup \mathcal{T}) \in \cup \mathcal{T}$, otherwise adding $f(\cup \mathcal{T})$ to $\mathcal{T}$ we would obtain again a bigger widening. So, the exemption of $\cup \mathcal{T}$ in the 2'.) , 3'.) rules are immaterial, in the case of $\mathcal{T}$. So $\mathcal{T}$ is a stage complete set and this was shown without using 4.).

Even more surprising is what we promised at the beginning of introducing the widenings, namely that $\mathcal{T}$ is actually $\mathcal{F}$, the minimal stage complete set.

This means that we obtained an alternative definition of $\mathcal{F}$ too, namely as 1.) , 2.) , 3.) , 4.). Unfortunately, this would have the same fault as the intersection of 1.) , 2.) , 3.) , that is we can't guarantee that there is any set satisfying it.
So what's the role of 4.) ? Well, it remained hidden because we made a big error.
We showed that $\mathcal{T}$ satisfying 1.) , 2'.) , 3'.) already implied 1) , 2.) , 3.) , which was trivial indeed. But, $\mathcal{T}$ satisfying 1.) , 2'.) , 3'.) is not trivial at all. In other words, the inheritance of 1.) , 2'.) , 3'.) from $\mathcal{W}$ widenings to their total, is the crucial step that requires 4.) which itself will inherit too.
We can see the importance of these more vividly from the followings:
Even though both the minimal or alternative 1.) , 2.) , 3.) , 4.) definition of $\mathcal{F}$ is imperfect without the existence guaranteeing $\mathcal{T}$ total, the question, why the minimal 1.) , 2.) , 3.) complete set automatically satisfies 4.) , or in reverse why 4.) implies minimality is meaningful in itself. The inheritance of 1.) , 2'.) , 3'.) , 4.) explains these two mysteries.
We can simply regard as special subsets the widenings in any stage complete z.
The simplest {{s}} widening shows that there are such.
We claim that the total of these must be the whole $\mathcal{T}$.
The cause of this, is the same that causes the inheritance of widenings, namely that they are beginnings or continuations of each other. Indeed then if the total of the widenings in a z were not $\mathcal{T}$, it had to be a beginning of this and thus continuable in z, contradicting that we collected all widenings.
Now, $\mathcal{T}$ satisfies 1.) , 2.) , 3.) so it is the minimal and automatically satisfies 4.) too.
But also any 1.) , 2.) , 3.) , 4.) satisfying z is a widening as well so can only be $\mathcal{T}$.
Otherwise it were a beginning of it and could only satisfy 2'.) or 3'.) not 2.) and 3.)

A more organized or I should say, well intended formalist way of introducing, the F inductive collection would go as follows:

1.) For every f function and s start, we have an intuitive spread or widening of the {s} start stage into an F final stage by adding the f (S) values to the S stages.
So we form S + f (S) as new stage but also form the union of earlier stages.
For the F final stage of course, either f is not defined on F, or f (F) ∈ F.

2.) It's impossible to tell if a set is this F just by looking at its elements
The collection melts in. But the collected stages, that is those special subsets of F that were used to reach F can be described. Namely, for the z set of these:
   1.) {s} ∈ z .
   2.) [S ∈ z ∧ f is defined on S] → S + f (S) ∈ z .
   3.) $\mathcal{S}$ ⊆ z → ∪$\mathcal{S}$ ∈ z .

3.) Unfortunately, these 1.) , 2.) , 3.) rules allow other rubbish too.
So such z could merely be called stage complete. The solution is to regard the smallest or minimal stage complete set, which can be produced as the

intersection of all stage complete sets. This then is the true $\mathcal{F}$ collection of the stages: $\mathcal{F} = \cap \{z \;;\; 1.) , 2.) , 3.) \}$

The final F stage of course is $\cup \mathcal{F}$, the widest element in $\mathcal{F}$.

4.) The logical question is whether we could add a 4.) rule to 1.) , 2.) , 3.) so that it would narrow a stage complete z set into $\mathcal{F}$, without intersection.
There is such rule 4.) = **. . .**
So we have to prove that a set satisfying 1.) , 2.) , 3.) , 4.) is the minimal satisfying 1.) , 2.) , 3.)

5.) The proof is based on a third concept, above the elements that we pick and the stages that we reach. Namely, we introduce the widenings, that is sets of stages, up to a stage. These will be denoted with $\mathcal{W}$ and the rules for them are:

1.) $\{s\} \in \mathcal{W}$.
2'.) $[\, S \in \mathcal{W} \;\wedge\;$ f is defined on $S \;\wedge\; S \neq \cup \mathcal{W}\,] \rightarrow S + f(S) \in z$.
3'.) $[\, \mathcal{S} \subseteq \mathcal{W} \;\wedge\; \cup \mathcal{S} \neq \cup \mathcal{W}\,] \rightarrow \cup \mathcal{S} \in \mathcal{W}$.
4.) **. . .**

As we see, 1.) and 4.) was kept exactly but 2.) and 3.) were changed so that f increasing doesn't have to be applied to the widest element in $\mathcal{W}$ and total union doesn't have to be added either.
This doesn't mean that we can not increase the $\mathcal{W}$ widenings the same way.

6.) The basic two results about widenings are that:
For any two of them one is a beginning of the other and the combining of any widenings is again one.
In particular the $\mathcal{T}$ combining of all widenings will clearly satisfy not only 1.) , 2'.) , 3'.) , 4.) but 1.) , 2.) , 3.) , 4. On the other hand any widening that is not $\mathcal{T}$ will be a beginning of it and thus continuable by f or union.
Then for any stage complete that is 1.) , 2.) , 3.) satisfying z, we can regard all widenings that are subsets of z. The total of these is then a widening which actually must be $\mathcal{T}$. Indeed if it weren't then it would be continuable by f or union. Since z is stage complete this could be done in z and this would contradict that we collected all widenings in z.
Thus $\mathcal{T}$ is subset of any z and $\mathcal{T}$ itself is a z, so it is the minimal and satisfies 4.) automatically. But also, any 1.) , 2.) , 3.) , 4.) satisfying z is a widening.
It can't be a beginning of $\mathcal{T}$, so it has to be the whole.
So in reverse too 4.) implies minimality.

7.) The use of $\mathcal{T}$ has an other surprising importance, namely, it is the only way we can prove the existence of $\mathcal{F}$. Indeed, neither the minimal 1.) , 2.) , 3.) satisfying z or a 1.) , 2.) , 3.) , 4.) satisfying can be shown directly, because we can't show any other example of a z satisfying 1.) , 2.) , 3.) .

I don't like this tidied up reasoning, because it hides the problems like all Formalism.
But as a recap it was harmless, maybe even useful.
Now its time to reveal the mystery 4.) property of $\mathcal{W}$ widenings:

4.) $S_1 , S_2 \in \mathcal{W} \rightarrow S_1 \subseteq S_2 \lor S_2 \subseteq S_1$

That simple? Well almost.
First lets observe why this should exclude rubbish intuitively.
The widenings grow by stages that have either a single new element, the f values or by the union of the earlier ones. Thus, we can't push any stage in between them.
Inside the smallest {s} stage, there is no room either, since it has only one element, so has no real subset. Finally, a widest rubbish can't occur either in a $\mathcal{W}$ that will be continued, because then the exemption to contain S + f (S) is only for the widest S in $\mathcal{W}$, so if it were a rubbish, then $\mathcal{W}$ would have to contain every possible stage. This at once shows that we have a problem:
Indeed, if $\mathcal{W}$ is the $\mathcal{T}$ total, then 4.) doesn't exclude a wider rubbish. In fact, if f is defined on it, we can continue it to include more. So in short, a restarted new widening can happen by f. Similarly, 1.) , 2.) , 3.) , 4.) wouldn't define merely the spread of {s} into F by f , but would also allow a continuing or even, many continuing new full spreads:

$\{s\} \subset \ldots\ldots F_0 \subset S^1 \subset \ldots\ldots F_1 \subset S^2 \subset \ldots F_2 \subset S^3 \subset \ldots$

To exclude such restarting can be achieved in two ways. Either we claim that the last element in $\mathcal{W}$ is the only one where f is not defined or gives an element:

$[\, f \text{ is not defined on } S \lor f(S) \in S \,] \rightarrow S = \cup \mathcal{W}$,

or we claim that {s} is the only element that is not ∪ of the earliers or f increase of the previous. To express this is more complicated, but more useful too, because it will lead us to the crucial application of 4.) to prove the inheritances.
The stages in $\mathcal{W}$ before S can be denoted as $\mathcal{W}$(S), this is the widening upto S and could also be called the history of S: $\mathcal{W}(S) : \{ z \; ; \; z \in \mathcal{W} \land z \subset S \}$
With this, the mentioned uniqueness of {s} is:

$S \neq \{s\} \rightarrow [\, S = \cup \mathcal{W}(S) \lor S = \cup \mathcal{W}(S) + f(\cup \mathcal{W}(S)) \,]$

Returning to the use of $\mathcal{W}$(S) for the main proof:

It's easy to see that $\mathcal{W}$(S) is a widening, but more importantly, we feel that this history of S is determined by f and s , not by $\mathcal{W}$. In short, $\mathcal{W}_1(S) = \mathcal{W}_2(S)$, so the $\mathcal{W}$ attachment is unimportant. Well that's the whole point of our proof.

We'll play with words. We know that $\mathcal{W}$ is immaterial, but temporarily assume a potential difference to show that it doesn't exist. This way we can move from the difficult question of what are the common stages of two $\mathcal{W}_1$, $\mathcal{W}_2$ widenings, to the much easier: What are the common histories in them?
But the trick is to ask an even more obvious question:
What are the different histories in one single $\mathcal{W}$?

One thing is obvious, due to 4.), namely that for any two $\mathcal{W}(S_1)$, $\mathcal{W}(S_2)$, one must be containing the other. An other way of saying this is that the wider is a continuation of the other. Thus for any number of histories they also continue each other and so their total combined should be either a history of $\mathcal{W}$ or the full $\mathcal{W}$ itself. This claim can be exactly proved too:

If every S element of $\mathcal{W}$ is element of a history, then the total of the histories will cover $\mathcal{W}$. If there is an S_0 element in $\mathcal{W}$ that is not covered by any of the histories, but all smaller than S_0 sets in $\mathcal{W}$ are, then $\mathcal{W}(S_0)$ is the total of the histories. Indeed, all S elements of $\mathcal{W}(S_0)$ are covered, so $\mathcal{W}(S_0)$ is covered too. But also any of the histories must be subset of $\mathcal{W}(S_0)$ or otherwise it would have to contain a wider element and thus cover S_0.

The only task is to show that such S_0 exists. It is quite simply, either the union of the $\mathcal{S}$ set of S elements in $\mathcal{W}$ that are covered by the histories, or if this $\cup\mathcal{S}$ is also covered, then $\cup\mathcal{S} + f(\cup\mathcal{S})$. To put it in an opposite way. If there was widest covered P, then $S_0 = P + f(P)$ or if there wasn't, then their total is S_0.

With this result, the same questions can be asked for histories that are not necessarily in a single $\mathcal{W}$. And yet, the same answers are true:

1.) For any two $\mathcal{W}_1$ and $\mathcal{W}_2$ widenings, one is history of the other.

2.) Any collection of widenings combined, is also a widening.

For the first, the trick is to regard the common histories. For these common histories, we can apply our previous results in both $\mathcal{W}_1$ and $\mathcal{W}_2$. So, the total will be a history or the full $\mathcal{W}_1$ and also a history or the full $\mathcal{W}_2$. Now it can't be a history in both, because then continuing the total with $\cup$ or f, we could get a bigger common history. So the total must be history in one and the total in the other, which is exactly the claim.

The second claim goes by actually checking 1.), 2'.), 3'.), 4.) for the total.

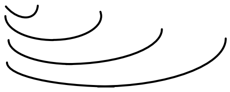
The rules all inherit trivially, except the first part of 4.), that is: $S_1 \subseteq S_2 \vee S_2 \subseteq S_1$

Now if S_1 came from $\mathcal{W}_1$ and S_2 from $\mathcal{W}_2$, then by the first claim, $\mathcal{W}_1$ or $\mathcal{W}_2$ is history of the other. Thus, S_1 and S_2 are both in one of them, so it is already true there.

Thus, the $\mathcal{T}$ total of all widenings is also a widening and it produces $\mathcal{F}$, the full set of stages. Then, the final stage is $F = \cup\mathcal{F} = \cup\mathcal{T} = \cup\cup\{z\ ;\ z = \text{widening}\}$.

The tricky verbal arguments we used to prove visually obvious claims raises the question if some verbal forms could be misleading. Obviously, the precise logically built forms can't be contradictory, but we always talk about what we see, in fact, our thinking is continually going along verbal forms. Some people even simplified thinking as silent talking to ourselves. This is false. But it's also false to claim that thinking is totally visual. The concept of widening seems to be very visual. The verbal opposite, "narrowing", jumps in our mind naturally, and we tend to believe that it's just a matter of saying whether we call a set of stages widening or narrowing. Going from smaller to bigger, it's a widening, but going backwards is a narrowing. This is a very convincing argument and it is true in some sense, but not in the sense we mean it by mistake. Indeed, widening in our concrete form was building wider and wider sets and after, infinite many, simply again regarding a wider. The elemental $f(S)$ increase and the $\cup$ to form the new wider one after infinite many, was incidental. We can start with any S_1 set, then wider and wider S_2, S_3, . . . and after all these, find an even wider S_ω. Then $S_{\omega+1}$ and so on. The same way we can start with S_1, find a narrower S_2 and so on. The only difference, is that here with narrowings, we feel a concern about the sets becoming too small, that is empty out, to stop possible continuation. With these visualizations, the reversal or sameness of widening and narrowing, is totally false. To see this, we should simply imagine the natural numbers as wider and wider sets of the beginning ones:

1 2 3 4 5 6



$\{1\} \subset \{1,2\} \subset \{1,2,3\} \subset \ldots\ldots \{1,2,3,\ldots\}$

Starting with this widest total $\{1,2,3,\ldots\}$ we don't have a next narrower. So, we simply can't go backwards in our widening. Of course, it's not impossible to find a next narrower in the total, for example, we could leave out 1. Then 2 and so on:

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

As we see, now we run out of elements.
What's more, this whole problem is not due to the use of sets and their containings. Indeed, we can regard the natural numbers themselves as an ordered set:

$1 < 2 < 3 < \ldots$

The backwards of this is confusing already, we could mean by it, the actual backwards of the same elements, but also a backwards by opposite as:

$\ldots . < -3 < -2 < -1$

This ordering of the negatives is different from the first.
So it seems the problem was that we only envisioned forward goings, both as widenings and as narrowings. And the forward of one is not backwards of the other.
These lamentations show the two main direction we can go from our functional induction or widening. The start of one line of generalization is just dropping the f function and s start value. That is, using:

$S_1 \subset S_2 \subset S_3 \subset \ldots S_\omega \subset S_{\omega+1} \subset \ldots$

Since this is not elemental by f and $\cup$, here between the stages, we could put new ones. Remember, for us, the avoidance of this was important, so that 4.) could avoid any rubbish. That rule, $S_1 \subseteq S_2 \vee S_2 \subseteq S_1$ is still a reasonable one. In fact, it expresses the obvious restriction that the stages are "ordered" by $\subset$. This then should be the only assumption, the framework of general widenings, in which we could specialize the forward order, that we called widening before.
The crucial concept we used in our proofs was the history. This seems to be perfect for any ordering. After all, $\mathcal{W}$(S) is merely the collection of all elements that are narrower than S. But this case is misleading. It gives the false impression that history must belong to a next element in $\mathcal{W}$, which was only true, because we went for such forward order. A more general concept of history that could be called only beginning,

can be defined as a set of elements that are contained in all the others, that are not in the set:

$$\underbrace{\subset \quad \subset \, . \, . \, . \, . \subset \quad . \, . \, . \, . \, . \, . \, . \, . \, . \, .}_{\mathcal{B}} \qquad \underbrace{. \, . \, . \, . \, . \, . \quad \subset \quad . \, . \, . \subset . \, . \, . \, .}_{\mathcal{E}}$$

The left half is the beginning, the right one is the end of the full $\mathcal{W}$.
So, $\mathcal{B}$ "$\subset$" $\mathcal{E}$ means that all elements of $\mathcal{B}$ are subsets of all elements of $\mathcal{E}$.
This seems to be stronger than the earlier requirement, that just one single S contained all the $\mathcal{B}$ elements, but it's not. Indeed, the later elements are wider anyway, so the singling out of S was stronger than using $\mathcal{E}$.
In this new more general concept of $\mathcal{B}$ and $\mathcal{E}$, we can have four scenarios:

$\mathcal{B}$ has widest element and $\mathcal{E}$ has narrowest element:	·······● ●·······
$\mathcal{B}$ has widest element but $\mathcal{E}$ has no narrowest:	·······● ·······
$\mathcal{B}$ has no widest element and $\mathcal{E}$ has narrowest:	······· ●·······
$\mathcal{B}$ has no widest element and $\mathcal{E}$ has no narrowest:	······· ·······

Only two of these mean that we had next S after $\mathcal{B}$.
So we can now define the special forward or well-widening sets that we considered before. These are the ones where every $\mathcal{E}$ has a minimal narrowest element.
It's strange that instead of the beginning or history, now the end became the crucial.
But we can put it in an other way which emphasizes $\mathcal{B}$. Namely, for every $\mathcal{B}$ there has to be an S, so that $\mathcal{B} = \mathcal{W}$ (S) . In short, every beginning is a history of an element. Just as in our earlier rules, the lucky generalization from beginning or history unions to any $\mathcal{S}$ subset union worked beautifully, now the well-widening can also be generalized to claiming that not just the ends, but in fact, any $\mathcal{S}$ subset has to have a first or narrowest. Indeed, regarding those elements that are before any $\mathcal{S}$ element, and those that are not, we get a $\mathcal{B}$, $\mathcal{E}$ division. Then the minimal of this $\mathcal{E}$ is also minimal of $\mathcal{S}$, so it must exist. But most amazingly, there is an opposite, seemingly much weaker condition, that already guarantees well-widening too. All we need to claim is that there is no backwards infinite sequence: $\dots \; S_3 \subset S_2 \subset S_1$ anywhere "in" $\mathcal{W}$. This "in" means that we can't pick such elements in $\mathcal{W}$.
Clearly, if there is such backward sequence in $\mathcal{W}$ then it is an $\mathcal{S}$ subset contradicting our previous criteria of always having minimal. But the more surprising side is that not having such sequence, we can't have any $\mathcal{S}$ either without minimal element. Or in reverse, if there is an $\mathcal{S}$ without minimal element, then there is a sequence too. The proof is quite "explicit". Lets pick an S_1 element from $\mathcal{S}$.

The elements before S_1 that is, $\mathcal{S}(S_1)$ must be infinite otherwise $\mathcal{S}$ had a minimal element. Thus, we can pick a narrower S_2 . Then $\mathcal{S}$ (S_2) is infinite again, so we can pick a next, and so on, we can easily pick a backwards, that is narrowing sequence.
This "no backwards sequence" definition of well-widening has an amazing negative form too. Indeed, not having such sequence, means that picking earlier, that is narrower stages can only happen finite many times. So what we said earlier about the emptying out is exactly happening in a well-widening. Every narrowing empties out in finite many steps.
Regarding arbitrary $\mathcal{S}$ set of S sets, the new questions were whether $\mathcal{S}$ contains widening or well-widening subsets that can not be extended anymore. Such maximal subsets of $\mathcal{S}$, clearly correspond to the total growth of an f from s. But these are less intuitive and more abstract.
The other even higher abstraction is to regard ordering of sets artificially, that is not by the $\subset$ relation. It's easy to see that any set with a $<$ ordering relation that:

1.) $S \not< S$
2.) $S_1 < S_2 \ \nabla \ S_2 < S_1$ Here ∇ means either-or.
3.) $[S_1 < S_2 \ \wedge \ S_2 < S_3] \rightarrow S_1 < S_3$

will give exactly the same possibilities as subsets.
In fact, regarding the beginnings, we get a corresponding widening.
So instead of well-widenings, the well-ordering of sets became accepted as the prototype of induction.
The well ordered sets have the same visual paradoxes as the well-widenings.
We can't have backwards infinite sub sequence, that is picking elements backwards, we can only do that finite many times before reaching the very first element.
The f , s spread is approached quite differently with well ordered sets:
We can still emphasize that f must give new elements for already achieved sets.
But now we claim without any apparent reason or "out of the blue" that the elements must be in a $<$ relation, which is well ordering.
This is to ensure that there is always a next one.
Then such well-ordering is simply an f , s well-ordering if:

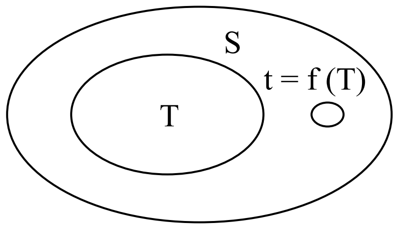
1.) s is the first element
2.) After any B beginning, the next is f (B).

There are obviously such f , s well-orderings, because $\{s\}$ is such.
The next is $\{s , f(s)\}$ with the $s < f(s)$ order. And so on.
This seemingly much smoother approach hides the whole problem of automatic growth, combinings, everything. In fact, it is directly claiming that the F full collection or stage is simply the longest f , s well-ordering. Of course, the precision is there, so the proof will be given that all f , s well-orderings melt into each other.
Namely, the two main claims are again:

1.) They are all beginnings of each other.
2.) Their union is an f , s well-ordering too.

Even the proof is the same, that is first these are established inside a single f , s well-ordering and then, that can be used in the heuristic way. That is, regarding the longest

common beginnings for arbitrary two f , s well-orderings. In fact, these arguments now split into the parts about orderings, well-orderings and f , s well-orderings, so in a sense, the arguments are clearer. But all this is misleading. We simply started with a lot of strong assumptions that in the end, pay off. And that's not all. The whole f , s induction is not even mentioned anymore. This is so, because it is applied to only one particular case.
Just as at the samples, the special {0 , 1} valued ones gave an amazing claim about the subsets of an S set, here too, there is a particular f that uses the subsets of S.
For every T subset of S, we can pick a t element, so that f (T) = t.
Starting from an s_1 element, then $\{s_1\}$ is a subset so $f\{s_1\}$ is a value, that is an element s_2. We can add this to $\{s_1\}$, that is regard $\{s_1 , s_2\}$. Then, f again gives a value and so on. The same vision happens as before, except that now we are remaining in S. If, f (T) $\in$ T, then the collection stops. This, by the way, must happen sooner or later, because for the full S, no matter what element we pick, it will be an element of S. Is it possible that the collection only stops at the full S?
Yes, it's easy to do. Namely, for every T real subset of S, we have to choose the f (T) value from S – T, that is outside of T.

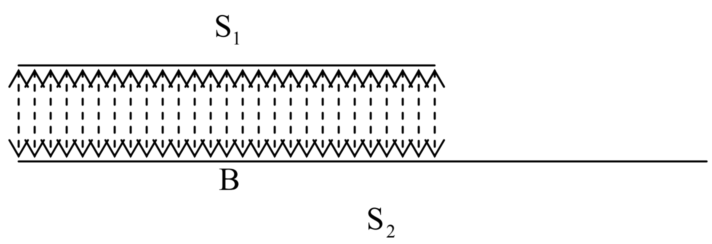


Observe that picking an outside t for every T real subset, and the start from any chosen s_1 element, the f will run its course. It will grow upto S, where it stops, or rather remains the same. So the F total growth or widened stage is S itself.
We might even say that this is the most stupid application. We got a total that we already knew it should be. But observe the particularness of the growth or widening is there too. Not all subsets appear in this widening. Not all possible f orderings are used. Which ones are is determined by the particular pick of the s_1 start.
In fact, the evil mind of Formalism can even hide this feature by accepting the empty set as a subset and then $f(\varnothing) = s_1$ is the start. Then, the run of f looks like a fix unique thing. The real result that is emphasized instead of the collection is the ordering of the S elements. Not even the f , s_1 well-ordering or f , $\varnothing$ well ordering, just the sheer fact that S became well-ordered. But this seems even contradictory. After all, the whole idea of f induction started with the assumptions of well-ordering and then specializing it as f well-orderings. But there is no real contradiction. Remember that we regarded longer and longer well orderings and f well-orderings. The S set having some well-orderings in it is obvious, because one element or two, and so on, are clearly such. But to claim that the picking can continue to the full S set, that is, the full S can be well-ordered, is now the big result. It is intuitive, but not an explicit collection. With the actual proof, it is reduced to explicit collection, namely, to the collections of continuations of the f well-ordered subsets.
But this side is not emphasized today. Instead the following:
What was the assumption used in achieving the well-ordering of S? Clearly, the existence of f that picks for every real subset an outside element. This f is not a timely process, it claims a space like feature. The possibilities of all S – T subsets are immaterial. Since they all have elements, we can pick one from each simultaneously.

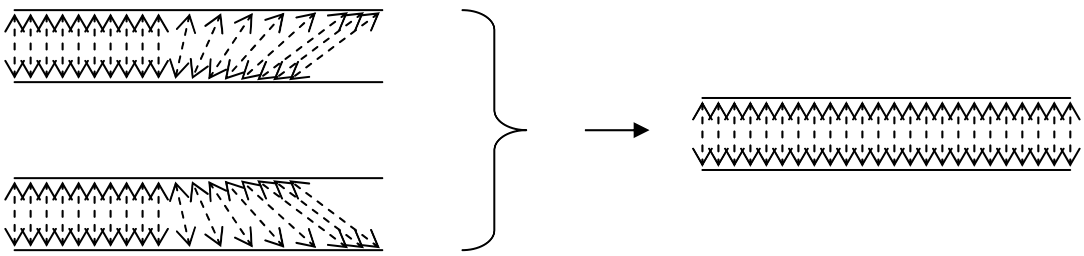
But observe, that this subset application of the f induction is different from what we did at the samples for $\{0 , 1\}$. There, we went from undetermined pickings to well determined concrete ones. Here we do oppositely. So while there, the problem of samples or simultaneous choices was left behind, here it is unavoidable. The existence of sample or choice function is so obvious, that it almost seems as a Logic step.
But it's not. Picking one element from a set, is indeed a Logic step, because if something exists, we can introduce a new name or variable for it. But to do this for infinite many as a collection, is beyond Logic. It was Zermelo who finally realized that we were dealing with a new axiom. In real arguments we intuitively use, not only this existence of samples, but the timely version of the self determined F too.
For example, when we proved that in an ordered set, an $\mathcal{S}$ subset without first element, implies a backward sequence, we did exactly this. We picked S_1 then S_2 and so on. The precise "timeless" way is to pick for every $S \in \mathcal{S}$ an earlier fix element, $f(S)$. Then, starting from an S_1, f will itself grow into:
S_1 , $S_2 = f(S_1)$, $S_3 = f(S_2)$, . . . This growth can be made explicit from f.
And the existence of f itself is the Axiom of Choice.
Among the lot of nonsense opposition to this axiom, the whole fundamental question of Inductive Collection was pushed aside. Both sides, the anti and pro Set Theorists, left common sense behind. A perfect example of how things changed is the following:
Back in Hungary, when I fell in love with Set Theory, the only thing in Hungarian was Kalmar's university text book. It was a deep book, in spite of having Stalin quotes in its introduction. The well-ordering theorem went on for pages and pages. I couldn't digest it at all. A few years later, I was waiting in Rome for my US visa and bought Paul Cohen's book. That was the first English book I read, in fact I practically learnt English from it. Here, the well-ordering theorem took three lines. Cohen wasn't a Formalist, yet this proof was obviously a shallow skeleton. Of course, his goal was to explain his new results of forcing. That has been since also polished to the bone.

5. The Continuum Hypothesis

Cantor was obsessed with equivalence. After all, that's what opened up sets to him. The comparison of size is important, but not the main feature of sets. Sets became the structures and models, the reality of all math. The reduction of inductive collection to mere well-ordering of sets also reflects this obsession. Why was well-ordering so important? As we remember, the crucial element of our proofs was that all widenings are merely continuations of each other. One is always a history of the other. This was of course for a fix f , s. With abstract well-orderings, the same claim is that for any two well-ordered sets, one is beginning of the other. This of course can not be true, because the two sets have totally different elements. What we really mean is that one is similar to a beginning of the other. So the intuition of the isomorphism principle is applied. This means that we can pair the elements so that the order is exactly the same.

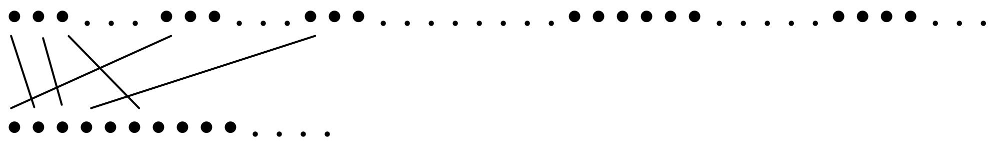


Now this at once means that S_1 is not only similar to the B beginning of S_2, but automatically equivalent to it too. In short, any two sets can be compared.
Of course, this equivalence doesn't mean $S_1 < S_2$ yet, it only means $S_1 \leq S_2$.
Indeed, an infinite set is always equivalent to some of its own subsets too.
The added extra task to establish $S_1 < S_2$ is to prove that $S_1 \not\sim S_2$, that is S_1 is not equivalent to S_2. So being smaller or bigger than an other set is always a dual process. One obvious is to establish that one is equivalent to a subset of the other.
The other really hard negative process is to prove that the reverse is impossible.
With finite sets, the first easy part at once settles the second. If a pairing of boys and girls in a ballroom leaves some boys without pairs, then we not only know that there are at least as many boys as girls, but that there are more. In other words, any new pairing will lead to the same situation of boys being leftover.
With infinite sets, one pairing does not exclude a possible opposite result. The tiny detail of producing an actual equivalence from opposite sub-equivalences, that is:



was explicitly shown by the Bernstein Equivalence Theorem. Thus, proving that Cantor's concept of smaller bigger is logical in this sense too.
The well-ordering of sets, not only proves that the easy first part of the comparison can be done for all sets, but raises some hope to go further. Indeed, for example we saw that the decimals are a bigger set than a sequence. So if we well-order the decimals as a D set, then a sequence can only be equivalent to some beginning of D:

A D well-ordering of all decimals:



This seems confusing first, because obviously the sequence is only similar to the initial first sequence in D. That's true, but we can use the single sequence in a non similar, only equivalent way. For example, we could use every second member to match the first sequence on the top, and then use the other elements to go further.
But smarter and smarter stretchings of a sequence can only go so far.
Namely, upto the first non sequencable beginning in D, that we denote as ω_1.
This notation is used because ω denotes a single sequence.
We definitely know that there has to be such unreachable point. But it might only be exactly the total D. So, the other question is what beginnings of D can be equivalent, that is stretched to the total D. If there are such beginnings, then they are also well-orderings of the decimals and so D was not the minimal well-ordering.

This is the most likely scenario if we regard D as just a "random" well-ordering and then the minimal in D can be denoted as Ω:

A D well-ordering of all decimals:

• • • . . . • • • . . . • • • . . . . • • • • • • . . . . . . • • • • . . . . . .

ω_1 Ω

minimal well-ordering of D

first beginning unreachable by sequences

The same way as the stretchings of a sequence, that is ω defined ω_1, we can define newer and newer next minimal well-orderings of the infinites:

ω is the first infinite well-ordering.
ω_1 is the first non sequencable well-ordering, that is the first non equivalent to ω.
ω_2 is the first well-ordering not equivalent to ω_1.
ω_3 is the first well-ordering not equivalent to ω_2.
.
.

Ω is not directly related to these next infinites. It is merely the first well-ordering equivalent to the decimals. Thus we can only be sure that it is not ω.
Cantor's question was, where Ω is among $\omega_1, \omega_2, \ldots$
He thought $\Omega = \omega_1$. Gödel believed that $\Omega = \omega_2$.
Paul Cohen proved that the question is undecidable from the axioms of Set Theory and he believed that Ω is "beyond" all the successively obtainable ω-s.
Of course this "beyond" is only a philosophical, or worse, metaphysical view. Indeed, the statement that Ω is one of the ω-s, is a theorem of Set Theory by the well-ordering of all sets. So if the undecidability was meant to be, as Cohen and I too believe, then the real philosophical question is what this theorem that Ω is one of the ω-s, really means. It claims an existing one among the ω-s to be Ω.
So it means either a new twist on existence or being Ω.
I think it is about a new form of existence in general.
I also think, this is related to the above mentioned concept of random well-orderings.
The well-ordering theorem creates such, from the randomly picked outside elements for all subsets. But this random side is not exploited in the actual proof.
That's why it becomes in conflict with the totally different inductive infinites.
The amazing actual method of turning the randomly picked elements into an effective collection, that is the explicitness of the "growth" from a fix starting element is the never emphasized "fine print" of the whole axiom of choice "debate".
This is what hides a finer distinction to be axiomatized.
There is a "half baked" approach to random picks, that "successfully" refutes the Continuum Hypothesis. Chris Freiling introduced it as the Axiom of Symmetry.
I will show this, by first proving that an alternative claim in general is equivalent to having an infinity between two $C < D$ sets:

Let $C < D$ be sets where C stands for "cardinality" because we'll only use the size of C, and D stands for "domain" because we'll use functions on D. Namely, f functions that order to any element of D a subset of D having size at most C.
A property of these f functions can tell exactly if there is a B subset of D, having size "between" C and D :

$$\exists\, B \subset D\,,\ C < B < D \quad \leftrightarrow \quad \forall\, f\ \exists\, x, z\ ,\ x \notin f(z)\ ,\ z \notin f(x)$$

For the → direction: Let $f[B]$ denote the union of all $f(x)$ values for $x \in B$.
This is the same size as B itself, so less than D and thus $\exists\, z \in D\,,\ z \notin f[B]$.
$f(z)$ is at most C size, so $f(z) < B$. Thus $B - f(z) \neq \emptyset$, so $\exists\, x \in B - f(z)$.
Then obviously $x \notin f(z)$ but also $z \notin f[B] \supset f(x)$.
For the reverse, we prove its negative form:

$$\text{not } \exists\, B \subset D\,,\ C < B < D \quad \rightarrow \quad \exists\, f\ \forall\, x, z\,,\ x \in f(z) \text{ or } z \in f(x)$$

The condition means that D is the next infinity size after C. So D is equivalent to the stretchings of C. In fact, this set of stretchings is a minimal stretching of D itself.
Then let f order to every element in this minimal stretch the earlier elements as set.
These are maximum C size sets. And yet for any two x , z elements, whichever is later in the minimal stretch, will have the other in its f value.
This is the same paradox as the following:
Let one person have only finite many idols. Then among infinite many people it seems natural that there have to be two persons that none of them is idol of the other.
But the naturals are exactly a counterexample by every number having the smaller ones as idols. All naturals have only finite many smaller, yet for every two numbers, one is smaller than the other. So we have to be careful about what seems natural.
In a random pool of infinite many people, the finite many idols are indeed minute. But the naturals are not a natural infinite pool. They are the minimal well-ordering of the sequencable sets. In fact we don't have a natural pool for sequencable sets.
Quite on the contrary, we do have a natural pool for the continuum, namely an interval. Picking a point is then the same as landing a dart on the interval as target.
But such interval also has a minimal well-ordering, an unnatural set, that allows to define strange f functions contradicting random feelings about the natural pool.
Now the "half baked" idea was to accept the existence of x , z elements as axiom for any f function with D being an interval and C being a sequence.
Indeed, then we get that there is B infinity between ω and Ω.
The "justification" for the axiom is that not only picking two points randomly from an interval must give different ones, but even sequencable "variants" of them must be disjoint. Already the difference of the points is only meaningful with distinctions of pure and random existences, which we don't have yet. Plus the dragging in of the variants is totally artificial.