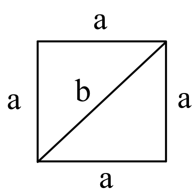
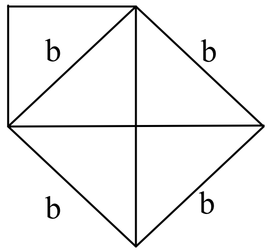
Behind The Well Ordering Theorem

More than two thousand years ago Greek mathematicians used concrete examples to show that there are a , b distances not commeasurable. This means that there are no m , n natural numbers that the m equal section of one would be the n equal section of the other.

That is $\frac{b}{m} \neq \frac{a}{n}$. A more practical form of this is $b \neq \frac{m}{n} a$ so using a as unit, the b is not fractionally calculable from a. In practical measuring we always use only such fractional distances. A carpenter goes down to millimeters, a cabinet maker maybe to tenth of this.
There is b distance that could never be exactly measured in such ways from a.
The most famous example of such b was $\sqrt{2}\,a$ which is by the way the diagonal of a square with the sides a:





By the Pythagoras Theorem $b^2 = a^2 + a^2 = 2\,a^2$ or we can see the same instantly by the second picture. So indeed $b = \sqrt{2}\,a$.

These pictures of course don't show why b couldn't be $\frac{m}{n}$ a.

This requires less obvious geometrical or algebraic tricks.

Indeed, it is not really a geometrical claim because $\sqrt{2}\,a \neq \frac{m}{n}\,a$ follows from $\sqrt{2} \neq \frac{m}{n}$

that is from $2 \neq \frac{m^2}{n^2}$ that is from $2\,n^2 \neq m^2$. And this is trivial from the unique prime factorizations of numbers because a square must have even many 2 factors.
So the left side as number has odd many 2 factors while the right has even.
Euclid produced much more complicated geometrical tricks to prove this same claim.
The Greeks were obviously obsessed with the whole division of distances and going down towards the individual points, which they regarded as a mystery. And it is still the same mystery!
What changed is that we tend to avoid head on attacking mysteries rather go for the "easy prey".
Find proofs for what we can, not search questions that seem to have no answers.
I am totally against this tendency and I think we gradually even forgot to ask questions!
But the practicality of exact proofs infected me too and without that I can not go for long.
An amazing fact is that though the non commeausurability of a and b can also be written as $m\,a \neq n\,b$, this non common multiplicity was not explored by the Greeks.
If we imagine the a and b distances repeatedly copied from a common start under each other like train carriage lengths, then this means never coinciding carriage ends.
Assuming this, if we regard the smallest differences at every end, we must have infinite many such possible minimal difference values.
Indeed, otherwise, that is with a difference repeating, shifting one line so the two same differences become zero, we would get a coinciding of the distances.

If there is coinciding then of course $m\,a = n\,b$ also means $\frac{a}{n} = \frac{b}{m} = u$ so a and b have this u as common unit and so measured in that, they are both natural numbers.
The possible end differences are all multiples of this u and so they are natural numbers too.
The possible differences appear again and again and their sums and differences must appear too.

This instantly proves that though the minimal differences up until the first coinciding seem quite irregular, they are in fact g, $2g$, $3g$, . . . arithmetically increasing values except in some crazy order. What's more, this g is actually the greatest common divider of a and b.
Euclid missed the common multiples as an easy visual road but he did use the minimal differences as remainders and even obtained an algorithm to calculate g.
Nowadays we call non commeasurability as the existence of irrational numbers because the a unit distance is automatically regarded as 1 and the $\frac{m}{n}$ fractions are called rational numbers.
This algebraization tendency was initiated by the introduction of the Arabic numerals.
These are actually Indian and hide the ingenious system of exponential number forms:

$$2014 = 2000 + 10 + 4 = 2\cdot1000 + 0\cdot100 + 1\cdot10 + 4\cdot1 = 2\cdot10^3 + 0\cdot10^2 + 1\cdot10^1 + 4\cdot10^0.$$

So to use the seemingly superficial 0 is the unavoidable trick in this shorter system.
One Pope claimed 0 to be the "devil's work" but the new system conquered Europe anyway.
The real magic of the new number system came by using it to measure all distances as infinite decimals. After the decimal point we decrease the 10 exponents to negatives which actually means division by ten, that is tenth hundredth, and so on fractions added with numerators that are the continuing digits.
On a lower level, at elementary schools, digit by digit calculation methods were taught for the basic operations.
Now with the use of calculators this might disappear but I hope not, for the sake of the most important fourth operation division. Here the method uses the remainders and brings down the next digit but after they are used up it continues to bring down zeros. So the process itself is always infinite. The remainders of course can only be as many kind as our divider, so we must encounter a repeated remainder. And since after a point we only bring down zeros, we must enter a cycle in our result too. We always get a periodic infinite decimal:

$$\frac{25}{14} = 25 : 14 = 1.7\underbrace{857142}_{\text{repeating period}}85\ldots$$

```
110
 120
   80
   100
     20
      60
       40
       120
```

Thus the non periodical infinite decimals are distances that are not fractions.
So we get the answer to the biggest headache of the Greeks as an almost trivial consequence of our vision of the decimals and our practical method of the division process. Plus observe that though we didn't get any concrete examples we got much more because we ourselves can create any non periodic infinite decimal or just accept any randomly created one as such and know for all these that they are not fractions. We can also feel how much more irrationals must be than rationals.
So what we have here is a most beautiful example of the so called "silver platters" that our accumulated human knowledge offers for anybody that simply happens to be lucky enough to be borne later in history. This raises some very important issues!
The first is that no matter how genius is someone at his own time, he can not escape the limitations of only possessing the earlier accumulated human knowledge. The accumulated simplifications can not be improvised by a single person! What's sad for of the individuals who were exceptional at their own times, is joyous for those who are average in a later time.
Everything comes to he who waits or rather everything comes to he who wants to know.
And yet, the motor are those who want to know all at their own time.

These are either doomed to blindness because they believe in fully knowing something or doomed to sadness because they realize their own limitations.
But the first crucial point behind all this is that if someone is ignorant and lazy to learn then even the accumulated silver platters are useless.
The second point is that this accumulated purification of human knowledge is in the hands of no one today. The exceptionals strive for new knowledge and the average is lured into ignorance.
To "teach", for the exceptional feels like a sacrifice, to stop the obsession of seeing new things.
In truth, clarifying what we see so that it can be transferred easier is the essence of seeing.
But this transfer is not simply displaying it, publish it to be digested by the other exceptionals.
Rather it should be explaining to the average who has only the will to understand.
And neither should this explaining be a lower version of what we see. Not a patronizing second grade version of analogies that present science "educators" perform in the Media.
And this missing intention of really telling what we see is coming from the simple fact that "seeing" is not the currency of our times. Our social currency is "action".
And the action of knowledge in math is proving. This "illness" of the present is not new.
Already Euclid was proving instead of explaining. Then Newton, Gauss, Einstein, Gödel were the four greatest minds who failed miserably to explain. The two mathematicians of these, Gauss and Gödel are the most obscene examples where derivation as single goal can lead.
Gauss kept his new visions about the non Euclidian Geometry in the drawer because there came no proof out of it! And yet the new proof was right there. Only the new vision was missing.
Having inner models of non Euclidian Geometries <u>proves</u> that the parallelity axiom is not derivable from the simpler axioms. And that was the actual initiator of the whole wandering towards new Geometries. The crucial hidden step is that sets as structures obey Logic.
Indeed, if the simple axioms would be enough to derive the parallelity axiom then in all models where those simple axioms are true the prallelity axiom, that is the single parallel should be true.
But we can see the inner models where the simple axioms are true yet unique parallelity is not true. So Logic can not derive this unique parallelity either. Vision flies ahead and doesn't care about the missing parts like how structures, models or logic should be defined exactly.
Vision can predict the correct consequences without the details. The exact proofs merely fill in the blanks. Analogies can not show the exact proofs. They are deceptions. But visions are the real skeleton of understanding. And the exceptionals know this very well. They are simply too weak to overcome the inherited taboos. To explain, to publish simply what we see is unfashionable! We must impress, we must prove! This is the poison of science.
Newton used as motto that "Physics beware of Metaphysics". Hegel reacted by claiming that he might as well said " Physics beware of thinking". The sad part is that Kant was between these two. He understood the importance of Newton, he went into it. And yet the recognition of understanding as the elemental action was missed. Kant's concern was abstract. How do we get this knowledge? And quite correctly he claimed that by a priori knowledge. But why do we get this knowledge so "badly" through only exceptionals, was not asked. Then he would have realized that the transcendence of knowledge hides the transcendence of all human conditions.
The distinction of exceptionals and average is the real taboo. And the only fix point we know is that everything that the exceptional can see the average can see too. So understanding is universal. The source that we pray to or gives the understanding as gift, is personal. The social reflection of this is a necessary lie. The transferability of understanding is the true substance of social existence. If we miss that, we will be punished at both ends. Those who lazy to learn and those who lazy to teach too.
In the twentieth century things sped up.
The long and tedious proof of Gödel for the fact that there has to be undecidable statements in Number Theory is driven by the single vision discovered by Turing, only a few years later that computability can not be complete. There are machines that can collect numbers or texts so that the missing, that is uncollected numbers or texts are not collectable by any machine at all.
The details of this delayed clarification are quite simple:
Just two years prior to Gödel's discovery, Pressburger proved that a simple addition Arithmetic is complete. So every statement is decidable from the known axioms. That, is either the statement or its negative is a derivable theorem. What makes multiplication so complex that

using it the situation becomes opposite, there are statements so that neither them nor their negative are theorems? The simple answer is that multiplication through the Chinese Remainder Theorem makes it possible that a unique c code number can be calculated for every number tuples. These are n-sequences of numbers. They are the remainders and the c code is obviously some very large value. The even more important consequence is that there is a reverse method too. So a single $T(x, y)$ number theoretical formula function can provide for every $x = c$ code and $y = i$ index value the i-th element of the c tuple.
The shocking consequences of this $T(c, i)$ sequence generator can be best illustrated by exponentiation.
In elementary school we define exponentiation as we did addition and multiplication.
Addition is repeated counting or 1 addition and multiplication is repeated addition.
Exponentiation is repeated multiplication and indeed:

$x\,x\,x \ldots x = z$ with having y many x-es gives: $x^y = z$.
The problem is of course that we can not use such dots in a strict language.
For the addition and multiplication either, and so they were not defined this way rather accepted as basic operations obeying the following axioms:

$$x + (y + 1) = (x + y) + 1$$

$$x \bullet 1 = x$$

$$x \bullet (y + 1) = x \bullet y + x$$

The first allows to obtain all correct addition cases by incrementing the second member.
The second defines multiplication for 1 second member.
The third allows to obtain all correct multiplication cases by incrementing again y.

Now you could say, that's nice so two new similar axioms can be used for $x^y = z$:

$$x^1 = x$$

$$x^{y+1} = x^y \bullet x$$

And indeed, we could do that but the whole point is that we don't have to.
We can do instead this:

$$x^y = z \text{ means } \exists\, c\ \forall\, i < y\ [\ T(c, 1) = x\ ,\ T(c, i) \bullet x = T(c, i+1)\ ,\ T(c, y) = z\ ]$$

$\exists$ means there is, $\forall$ means for every and these are proper part of any language and so this formula then indeed defines exponentiation without dots.
This would be a monstrosity to use as definition of exponentiation in schools but the point is that it is identical with that naïve definition and also with the above way of extending Number Theory. So containing multiplication is a big leap in itself and we can even realize that not only exponentiation but actually all ruled new operations can be defined inside, without accepting them as new basic concepts.
The derivations in this fixed Number Theory are also just some new operations definable by rules and so actually expressible as formulas if we can correspond numbers to all expressions in our language. And Gödel did exactly that. This also allows a listing of all derivations and a listing of all one variable P( ) properties. Then we can create a $G(x, y, z)$ formula which claims that the x-th P( ) property at the y value as statement is derivable by the z-th derivation. Then $\exists z\, G(x, y, z)$ simply means that the x-th property at the y value is a derivable statement.
Observe that $\neg \exists z\, G(x, x, z)$, that is a negative and common variabled formula obtained from $\exists z\, G$ is a P( ) property and so is actually a k-th one in our list.
Now the big question is what happens at $\neg \exists z\, G(k, k, z)$?

This is a statement and it says that the k-th property is not derivable at the k value.
If it were by an n-th derivation then $G(k, k, n)$ were true saying that $\neg \exists z\, G(k, k, z)$ is true so two contradictory claims were true. Thus $G(k, k, n)$ is definitely false for any n.
So $\exists z\, G(k, k, z)$ is false too and so $\neg \exists z\, G(k, k, z)$ is the truth.
But what about derivations? Amazingly, this truth is definitely not derivable if our theory is consistent that is if it has no contradictions. Indeed, a derivation of $\neg \exists z\, G(k, k, z)$ if it's the n-th, would imply a derivation of $G(k, k, n)$ by G's meaning but also by sheer logic $\neg \exists z\, G(k, k, z)$ would imply $\neg G(k, k, n)$ so we would have a contradiction.
This doesn't mean an undecidability yet because we could have as false theorem the derivation of the opposite that is $\exists z\, G(k, k, z)$. Would this lead to a contradiction? Not at all! It would only mean that we have a derivation of an existence and yet all n cases of z were false.
So the non derivability of $\exists z\, G(k, k, z)$ can be derived not from mere logic but from an added assumption that Number Theory should not have such existences as theorems.
If all cases of a property is false then an existence should not be derivable. This sounds logical but it isn't a logical step like avoiding contradiction in indirect arguments. Amazingly, this hides our much bigger assumption of "consistency", that no contradiction exists at all. Indeed, surprisingly to outsiders, in math one contradiction means that everything would be derivable.
Still, this assumed consistency is a much weaker condition than the non existence of false existences! Indeed, from a contradiction everything is derivable so having any non derivable statement implies consistency automatically.
A more tricky formula and argument that creates undecidable statement from mere consistency was discovered few months later by Rosser and it was important because this assumption of consistency is really the minimum. In fact, Gödel went into this direction and showed that such self consistency claims must always remain unprovable by a system if it is indeed consistent.
All this drove these new results toward language, expressability and so on.
Turing was the first to say no to all this nonsense! He saw that the real reason of having undecidable statements is that the derivations simply collect a subset of statements, the theorems so that the complement or leftover subset, the set of non-theorems is not collectable by any system of rules. If this is so then indeed the claim that every statement or its negative is derivable is impossible. Simply because after deriving the theorems we can just negate them and voila we would get a system that derives the complement set, the non theorems.
It's the complexity of the non-theorems as set, that makes it non derivable by any means.
This forces out the fact that then the negated theorems, the anti-theorems or derived falsities can not be identical with the non-theorems. Once you see this, you got the whole new point.
For me personally by the way, the coin dropped after reading a little book by Rosser.
But Turing had more in his mind! He also saw that this derivability by any system is actually a mechanicalness that will be a new field, computability. Computers of course became the present since then and this unfortunately can mislead our minds.
Mechanical systems were already existing earlier and using electronics is really not the point.
But even mechanicalness isn't. The best example is games. These have strict rules and indeed allow a very good demonstration of what lies behind Gödel's result.
The derivable situations in chess are all game points where two non cheating players can go.
This even parallels the seemingly so important free thinking in mathematics.
Just as good chess players think in some hidden world of tendencies and ideas, mathematicians do the same in the realities they project behind the mathematical relations. The use of axioms is just as unconscious for a mathematician as to obey the game rules for a good player.
The only big difference is that games unlike statements do not have a trivial negative.
The new negative that Turing brought out is not the negative of individual objects rather the complement of whole sets of objects. And this exists for games too! These are the situations not obtainable by game history. To put chess pieces on a board randomly, we get the amazing question what the hell is here? Was it achieved by a game or is it an impossible situation?
There are obvious clues! Like the bishop that remains on same color. So a phony set up is revealed at once by two bishops on same color. But aside from these trivialities, we see that while a true situation has a game history that proves it, the impossibility of a phony situation can only be proven by checking out all possible game histories up to infinity.

So the false game situations is a more complex set than the true ones.
The step of forgetting about creativity and free thinking was made already by Gödel when he realized that the possible derivations is a simple list. We can use the axioms and rules in any particular order and go through all derivations. Though we'll need some tricks even for that.
But stepping from free derivations to mechanical listings is not enough!
The mathematics of this was ready for Gödel as primitive recursive functions.
The word "recursive" hides actually both ends. The fact that when we derive theorems we use older theorems is a kind of recursiveness. But the primitive version merely steps through inductively from 1 all possible values consecutively. The common bad part is that we are still using numbers and operations. These are abstractions! Turing returned to elementary school!
The digital calculations are the real concrete computations with natural numbers.
Our instant reaction to this should be a shock! How can we regard these as a base of anything when these are dependant on exactly the operational meanings.
But we should forget about these particular methods because we aim for anything mechanical.
We should allow digital alterations by any rules. But keep our framework of rules still as simple as possible. So we stick to digit by digit alterations and stepping to next positions, just as the elementary school calculations do but we even restrict these to work on a single line.
First we might think that such methods can not possibly even accomplish the elementary school calculations, but they can! We simply have very slow "crawling" methods to do the same.
Every alteration that interrelates with digits far away must be done and then crawl back to the other digit then back again and so on.
Such crawling computer if were built as real, would require trillions of years to do anything that our modern computers do in a minute. But the point is that all that can be done on a modern computer, can be done on a single line of memory cells too.
The even more important point is that actually we have more than in any real computer.
Indeed, even though we lost time, we gained space! The infinite line is still infinite and so we can do things that no real finite memory machine could. All information of the universe could fit on an infinite line bit by bit.
This data alteration idea offers as almost natural consequence that the effective functions then should use the input data as variables and the altered data as values. But here we took a ballast with us from the past. The recursive or effective functions were actually ad hoc choices to collect numbers effectively. Derivation systems do collect objects directly and use no functions but this direction failed too. Only the equation systems for functions survived. We still don't see clearly why this obsession with functions had to rule effectivity prior to Turing but now it became free of this. So to regard the altered data as value is narrow and old fashioned.
The alteration is actually an examination of the input. It can use the whole line to create new numbers and compare those with our input. All this is a search that may not even end.
If our alteration is running for ever then there is no value. To tell if an input will be okay or not can only be by running the recognition alteration process. All this was already known for recursive functions as opposed to primitive recursive ones. But there this was accepted as an inconvenience. Namely, that only functions can define the wider effective collections that do not decide, merely recognize the collectable objects. These can be domains or range too but still annoyingly use functions.
Turing's alteration process opens a new door to collection! Forget the obtained values as function values! There is no function we aim at. We aim at recognizing the objects to be collected. So we collect all those inputs where our particular alteration process stops.
The free derivation of expressions was changed by Gödel to listing numbers and now this is changed to recognizing numbers.
Amazingly, we can also step back to the original wider field of expressions! Number alterations naturally offers itself as text alterations in general with any fix alphabet. Of course, numbers or even number tuples are automatically included.
That's the good side! The seemingly bad side is that we don't have the simple lists. In fact, it seems that these acceptable inputs where a method stops can not even be listed because by trying one we may be unlucky and stuck forever. To a total outsider even the fact if we could try out all texts as inputs is doubtful. For numbers we have the God given increasing values but

for texts how should we go. It is called the “length alphabetical” order. A dictionary is alphabetical but only works because the words are limited in length. If all sequences of “a”-s are allowed then we never get to “b”. The solution is to first list the 1 length texts then the 2 long ones and so on. The 1 long ones are simply our alphabet in alphabetical order. Then the 2 long ones have all pairs in alphabetical order and so on. All this still leaves the bigger problem of the mentioned failing at a tried input. To solve this we use “dovetailing”.
We start our alteration process on “a” the first possible input. We do a few steps and then step to the next “ b”. Here again we start and after few steps we return to a few more steps on “a” and then again on “ b”. then we step to “c” and from there again we step back to all earliers.
This way we get to all data and eventually the recognized ones will be recognized but in a totally ad hoc order.
Observe that the recognition of a data has nothing to do with how short is that data!
The Bible may be recognized as acceptable in a few steps while the word “God” could take trillions of steps. Simply because we may have to compare these with more complex things.
This at once implies that our dovetailing recognition will not be in increasing order at all.
Amazingly, it’s quite obvious that if by any method we have an increasing listing of some texts then both this list and its complement are acceptable collections by a method.
The first part is true for all listings. To get a recognition of a t input data we just have to wait until our listing pops out t. To identify a missing t one of course would require to wait for ever so this doesn’t work for the complement set. If however our listing is length increasing then we can wait until the first listed data appears that is longer than t. Then we can see if t was there just before. among the ones as long as t or not. If not we should accept t.
This acceptance will collect exactly the t-s not in our list.
An other way to look at this is to list all possible inputs length alphabetically and in that tick out the length increasing list members as they come. After any new length, the previous length ones that were not ticked out can be listed. This gives a length increasing list of the complement.

So the three big phases of Effectivity were “derivability by rules”, “listings by numberings” and finally “recognition by alteration sequences”. If we reveal that every concrete Turing alteration method is given by a table of rules that tells the cycles of alphabet changes, movements of the changer and changes of an inner state of the machine, then we may say: okay so the finite rule system is the essence. But a new twist entered by Turing too. While in derivational rules and game rules, the rules must be allowed to be more and more complicated that is be more and more in number, here with the use of inputs we can place all this complication into there.
We still need a table that tells how a fix special machine works but this then will be able to use part of all inputs as programs to imitate other machines. So the program part of the data will tell how to alter the rest. This table must be smart enough to use a program part of any data as program. If we have such a table then all the more and more complicated that is bigger and bigger tables can be avoided or rather incorporated into the arbitrary long data. The word I used above was imitation but we must be careful with this. A program runned alteration sequence is obviously totally different from how a simple data alteration goes by a big table. This second would be much faster than running back and forth to a program, but we already talked about how irrelevant time is for us.
We really must tell in what sense are our programs the imitations of tables.
And this sense is amazingly loose. We only claim that they should collect the same inputs.
In other words, stop or run for ever from same inputs accordingly. An other name for this is that they as collecting machines are merely variants of each other. A drastically opposite weakest version of input relation is if two machines only “share an input”. By this I mean that for a t input data both machines stop or both run forever. To share an input is almost obvious if we just chose two machines randomly. But here is a crucial question. What if a machine is such that it shares input with all other machines? The amazing triviality is then that this machine can not have a machine that would be its opposite, that is collecting exactly all those inputs that our machine left out. Indeed, opposite machines can not share an input.
Having an underivable complement meant for the set of theorems that the non theorems are not derivable and this implies that there are undecidable statements.

Now having underivable complement means being collected by a machine so that the collecting machine shares an input with all others. Turing realized that those fix U universal machines lie behind this that can imitate all others. Almost always this situation is oversimplified but unfortunately we have a real problem here. U uses two inputs, the p program and a real t data. It's easy to say that all program is actually data because it can be formulated by some alphabet. To show that data alterations can be done by a p is not that hard but then accepting that alphabet for t too, will mess up the running of our program. So we must code our program alphabet into the regular data alphabet. Then every p program can be regarded as s P obtained from regular s texts by a P coder. So (s P , t) U is using only the normal alphabet in both s and t. But then still the imitation from all single inputs is false.
A perfect imitation even for collection of same inputs is impossible by a U !
Luckily we don't need imitation! We only need sharing input with all machines.
So we start with a process of D duplication of our arbitrary t input into a new text t D that hides the t P program version of t too! This involves the rules of space on our line, how we should later use our program. In an abbreviated sense t D = (t P , t).
Simply copying the data next to itself with some separation is not enough.
Turing used the even and odd memory cells which simplifies the actions.
Observe that D is not a machine that may stop or run forever, rather an alterator that always gives an end result. We continue by a U* machine that executes U. So t DU* = (t P , t) U.
Only D U* is real machine or table. P and U were just the ideas behind
If p is a program for an M machine and t is an input that translates into this p then:
t M stopping or not exactly happens as t DU* stopping or not.
So the only important part was that every M machine should have a p that imitates M and every p has a t so that t P = p. And then the trivial result is that DU* will share input with all machines and so can not have a complementing machine.
This placing of the same input as program resembles the Gödel usage of same index and value into his G relation.
The proof by Turing that there is a machine that shares input with all others did not reprove what Gödel did. The crucial role of multiplication in all possible axiom systems can be seen as an analogue of universality but this does not avoid the Chinese Remainder Theorem and the rest.
And yet Turing went further because he showed a deeper reason for the undecidability than the particular statement. The theorems are a derivable or listable or recognizable set that has a non derivable or not listable or not recognizable complement. The particular statement then sneaks back to prove the universality and input sharing. But proofs are not the essence. It is the vision.
Because vision talks about the facts. The non theorems being such a complex set is the reality, the hard fact. How we prove this is just our human business that at present rules everything.
If anyone was aware of all this then it was Turing and yet he fell into the same trap as Gödel and Gauss before. Tried to convey his vision only through proofs. Of course, eventually the visions rule not the proofs and Turing was also accepted not due to what he proved rather what he offered as new vision. But this is the hidden untold part of history. The education system is at total loss and can not transfer what happened even at such recent events. And that's not all!
The biggest travesty of lie is about what took twenty years to surface after Turing.
The true depth of Turing's own baby, the program was missed by him. After the War even a first Bible of the new Meta Mathematics was written by Kleene. It missed the vision too.
Then the recognition, what programs can really do became a seemingly minor new theorem.
Usually just shoved into the old line of proofs as if it were just a previously missed detail.
Then gradually it became a "must", mentioned by all text books. The question that if this is so important yet so simple then how could it be missed by Turing and Kleene, is never raised.
It's a taboo because it shows with razor sharpness how futile is rampant formalism and worship of proofs alone. People like me who believe in that simple failings in our visions are that hold us from solving the biggest open problems are regarded as crazy.
But the real crazies are the "sober" ones because they falsify history! They don't know or deny what happened at Cantor and again at Turing. Or they admit some parts but believe that history won't repeat. Because now mathematics is different! It's team work! It's details not big recognitions! So only the future will prove me right! And history always repeats.

The Cantor Turing parallel is perfect and also the missing second deepening of concepts.
The missing deepening of Sets became the choice functions and the Well Ordering Theorem.
The missing above mentioned deepening of programs became revealed by Rice's Theorem.
This theorem opened a whole arsenal of effectively collectable texts that all must have complements effectively not collectable. And anybody can see this easily in minutes!
To prove it, I will now explain this theorem. It has three phases. A simple and amazing stronger claim that implies it. The trivial step of stating the theorem itself. And finally the mentioned consequence that implies the arsenal of examples.
The stronger claim is the impossibility of an E machine that actually doesn't seem so impossible at all. But it implies that any M should have a complementing machine so from Turing's result we can infer that E doesn't exist. The letter E stands for "empty" but not directly in three sense!
We don't want to talk about the p programs in their own alphabet, so again we obtain them as s P from regular s texts.
s P being a program for an M machine means that (s P , t) U imitates M but our strict condition was merely that they should collect the same inputs that is stop from exactly same ones: (s P , t) U stops ↔ t M stops. This imitated machine for any s will be denoted as (s) and the collected inputs for an M machine as [M] but [(s)] abbreviated as [s].
This is especially logical because we want to avoid the whole imitation and U.
The important point is merely that every t text has now a [t] set of text ordered to it too.
Being an empty machine should mean to collect nothing that is to stop from no input ever.
We use empty for the t texts in their program sense too, so it means not that t itself is empty rather that [t] is empty. The E machine will not be an empty machine rather an empty text collector. This the first misleading in the letter E. Secondly, we don't even mean an exact collection or recognition by stoppings. normal recognition means that we stop from these inputs and only from these. Now we allow a weak recognition, so stopping from these empty texts but maybe also stopping from others. This of course sounds crazy because a machine that stops from all inputs is at once a weak recognizer of anything including emptiness. So we need a third twist to avoid this triviality and we'll not merely exclude a fix s_o text but all variants of it.
This is meant again in the collection sense. So we can describe our E by:
[s] is empty that is s is empty → s E stops
[s] = [s_o] that is s is a variant of s_o → s E runs forever.
Our goal is to show how our E would imply a complement for any M so we use this M to define an A alterator. Of course, A depends on U and our chosen s_o too.
Imagine the following process: For any s , t pair of texts we start M from s and if it stops we start N from t. Knowing M and N as tables this were quite easy because the stopping from s is actually just reaching a state in M's table that has no continuing cycles.
We can alter this by adding new cycles and executing N from t. But we claim much more!
Namely, that all this is doable by an A alterator. So s A P is a program for every s that using t as input will be a variant of our process through U.
And we chose N as the machine determined by the s_o P program through U. So:

(s A P , t) U stops ↔ s M stops and (s_o P , t) U stops. Observe that:

s A is empty if s M runs forever and s A is a variant of s_o if s M stops.

So: s M stops → s A E runs forever and s M runs forever → s A E stops.

So: s A E stops ↔ s M runs forever. So A E is a complement machine of M.

Now comes the second phase, to claim something seemingly much stronger than our claim was above yet actually being a trivial consequence.

Suppose we have a T set of texts that is not an empty set so we have some texts inside T and is neither the full set of possible texts so we have some texts outside T too. Plus for every t in T if an other t' is a variant of t then t' is inside T too. So T is variant complete.
Then if this T set is collectable by a machine, the complement of T is not.
This is Rice's Theorem. And its proof is almost trivial now.

Indeed, if the complement $\neg T$ were collectible by a machine too, it would imply at once a forbidden E. Simply because first of all the complement of any variant complete set is trivially variant complete too, so $\neg T$ is variant complete too and the empty program variants must be in either T or $\neg T$. Thus the machine that collects this half would be an E.

Now comes the most amazing third phase, the application where we can make such T-s in abundance. All we need is some F finite property that we watch for in the runnings of the p programs as sequential collectors. So we use length alphabetical listing of all texts and in them apply the M machine determined by a fix p in dovetailing. The result is a spitting out of texts where we look for some F finitely observable occurrence. For example if we spit out numbers we can watch if a particular number say 10 appears. This is trivially an effective thing and so the collection of the p programs that will spit out a 10 is an effective T set.
Thus those programs as set that will not produce 10 are not collectable effectively!
You can not tell from the programs if they will miss to produce 10.
For some particular programs you may foretell this but not in general exactly.
In old fashioned number theoretical collections we always collect by relations and operations.
That is why there the complements are always effective. Like the composites or primes. Both collectable by machines. The secret is to run the numbers as programs! This is the new deeper vision of programs!

The corresponding deeper vision of sets are the choice functions! And these lead to the Well Ordering Theorem which is our actual goal later. That took even more than twenty years to surface.
But I will still ponder about other facts before I dwell into that mystery.
Both Cantor and Turing realized the beehive they opened by the concept of set and program. And yet they were unable to see the next clear application of these concepts.
A simply provable fact remained hidden from their eyes even though they were trying so hard to use what they saw. This is a deepest mystery. The narrow thing they did prove clearly was a "tunnel" connecting their new world with the old.
Cantor's tunnel is exactly where we started and now we return to, the irrational numbers.
The infinite decimals and the division process are the most beautiful silver platter that made the old fashioned Greek irrationality proofs obsolete.
Cantor's tunnel is actually a new silver platter making again obsolete something that some 19^{th} century mathematicians proved. It was again a meticulous concrete way to show that some distances are not belonging to a special class. An old special class wider than the rationals was the constructible distances inherited also from the Greeks but the characterization of this was finished by the beginning of the 19^{th} century. Gauss was a teenager when he made a breakthrough in this characterization and this pushed him to be a mathematician. But the class we come to is more algebraic, in fact these numbers are called Algebraic and are the roots of any equations that use a single x variable as unknown and some fix rationals. The allowed operations are of course the algebraic ones but x can not be in an exponent. As it turns out, these can be all reduced to simple power sums of x with whole numbers multiplied. We also call these as whole polynomials. They were also famous for possible whole solutions with more variables for a long time. With single unknown but allowing infinite decimals as solutions the solvability was also a long standing problem. Gauss was again the one who realized that his new complex numbers are the field that guarantees always roots. It's called the Fundamental Theorem Of Algebra. A side consequence is that we can have only maximum n roots if the highest x power is n. But the class of real numbers, that is the distances or infinite decimals as possible roots has nothing to do with this. It was believed almost with certainty that there has

to be decimals that can not be such roots and these non algebraic numbers were even named as transcendental, but the proof of their existence was a hard task. The situation was even messier because the two famous constants, the old Greek π and the newer Euler constant e were also believed to be transcendental. The proof for these came even later. So a smart but artificial decimal construction was all that proved the existence of non algebraic or transcendental numbers when Cantor stepped into the arena. His vision was this:
The collection of all possible decimals even just restricting them between 0 and 1 that is starting without whole part so starting with the decimal point is an incredible big collection.
These are not sequencable! Any sequence of such decimals must necessarily miss some.
So whatever class of special numbers could be sequenced must obviously also miss some too and so there must be numbers outside such special class.
This grand idea can even work for the now trivial irrationality and so it's educational to reprove that too. As it turns out then already here a not so trivial first part becomes that the special class, the fractions is a sequence. In fact, the whole Greek obsession was initiated by the fact that the fractions are very dense on the line. They can approach all distances just can not be themselves all of them. So to say that all these fractions even on a small interval are a mere sequence like the natural numbers is quite surprising. But that all of them on the infinite line are just a sequence, is almost a paradox. And yet the proof is so simple. We just use the earlier mentioned length alphabetical ordering idea. But what should be now the "length" of a fraction? The point was that within one length we had only finite many cases. So the length should be the total of the numerator and denominator because with one such value indeed we have only finite many cases and these can be listed increasingly. So the full list of the fractions is then:

$$\underset{2}{\frac{1}{1}},\ \underbrace{\frac{1}{2},\ \frac{2}{1}}_{3},\ \underbrace{\frac{1}{3},\ \frac{2}{2},\ \frac{3}{1}}_{4},\ \underbrace{\frac{1}{4},\ \frac{2}{3},\ \ldots}_{5}\ \ldots$$

The crucial general claim is of course that the decimals are not sequencable, or rather that any sequence of these misses some. So let one sequence be:

$D_1 = .6\,3\,0\,5\,9\,1\,0\,2\ \ldots$

$D_2 = .1\,5\,8\,3\,0\,4\,2\,6\ \ldots$

$D_3 = .2\,0\,9\,1\,6\,3\,0\,2\ \ldots$

.
.

Lets form the diagonal digits as a single diagonal decimal:

$D = .6\,5\,9\ \ldots$

This would be strange in our list but nothing forbids it to be there.
Lets alter every digit in D by any way, or to be more specific lets add to all digits 1 but for 9 this should mean to get 0. Then this D^+ decimal is:

$D^+ = .7\,6\,0\ \ldots$

This can not be D_1 because their first digit is different. It can not be D_2 because their second digit is different. And so on D^+ can not be any one in our list.
This is the point where your heart stops a beat and you are either shocked by how simple beauty just happened, or you are shocked with hate what nonsense was just presented to you.

But lets go further! According to Cantor, this same argument works for the Algebraic numbers because they are also a mere sequence. The “length” now should be the total of all absolute values of the wholes used in a polynomial. For example $3x^4 - 5x^3 + 2x^2 - 7x + 1 = 0$ has as total the $3 + 4 + 5 + 3 + 2 + 2 + 7 + 1 + 1$ value. For any such fix value we can manufacture an incredible number of equations especially considering the negatives too.
Still, this is only a finite number. So we can list all polynomial equations by increasing totals.
Then knowing that every n-th order such has maximum n roots, we can list the roots too.
It’s amazing that Kronecker the greatest nemesis of Cantor denied this ordering of the Algebraic numbers. Probably in his mind this last step of inserting the roots was a “no no” because these roots are very hard to determine. Today we are beyond that. Just because we don’t know something we can still think about it in terms of what we know about it.
I don’t want to excuse Kronecker in any form. The way he behaved personally against Cantor deserves only one adjective “scumbag”, but we must admit that this over thinking of object determinacy remained alive in the opposite camp too. The possible collection of all decimals still had the hidden assumption that these should only mean the ones that have some finite determinacy. This is apparent by a dilemma brought up by Brower and König.
The idea of randomly chosen digits and so the acceptance of totally undeterminable decimals as objects was not embraced yet.
Now we can come to the bigger world what this tunnel proof of Cantor wanted to introduce.
In this new world the infinites have different sizes!
The fact that infinites are big because we can steal from them and yet they remain infinite was well known. Even the more surprising fact that we can duplicate them without “loss”.
Galileo when formulated his falling law quite unfortunately using the odd consecutive falling distances, realized the “curious” fact that the odds and evens are actually “same sets”:

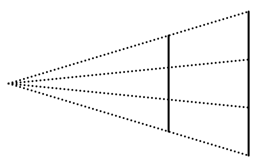
$$1, 2, 3, 4, 5, 6, 7, 8, \ldots = 1, 3, 5, 7, \ldots + 2, 4, 6, 8, \ldots$$

The simpler stealing of finite many elements, that is:

$$1, 2, 3, 4, 5, 6, 7, 8, \ldots = 6, 7, 8, \ldots + 1, 2, 3, 4, 5$$

Is somehow much less disturbing.
All these give the impression that though the infinity is deceptive, the actual collection of the objects is a fix amount.
I found my own paradox when I was young as follows:
Imagine beggars in a row having all a dollar in their cap. Now if I steal the dollar from the first but replace it from the second and that from the third and so on, then I gained a dollar and actually stole from nobody. So stealing infinitely can make free money.
No wonder I remained poor all my life.
The above forward stretchings of a sequence open amazing visions.
The simple finite stealing to put forward is failing after infinite many repetitions. We simple get back our original sequence. The Galilean half division seems to go much further but it will stop too. So it seems that stretchings forward have a limit of infinities.
The big confusion here is that these stertchings are actually all different states of a fix set so we feel that the infinity should be the same but we also feel a definite increase. Today we call these Well Orderings and they were crucial to clear something about infinites.
Cantor realized that a definition of equality among infinites should not rely upon these stretchings because they already regard only fix sets. The abstract concept of being equal in size should mean merely a one to one assignment of the elements called equivalence.
This was what we created between the naturals and the fractions by our sequencing of them and this was that we showed to be impossible between the naturals and the decimals by our anti diagonal construction.
The idea of stretchings is still relevant because all forward stretchings relate to each other by one being a beginning of the other.
So then this stretchings of the sets became relevant for two goals for the infinites.

One simply is that these stretchability jumps are actually the exactly increasing infinites and so we can ask for any two known different infinites like the decimals and the naturals how many jumps are between. Stretching the naturals as far as possible we get a total of these stretches as a well ordered set, a super stretch! But this can not be a stretch for the naturals because then we could get a new bigger one by stealing again. So this total stretch has to be non sequencable!
Thus we got a new method of bigger infinite and quite naturally it has to be the next after the infinity of the sequence, the smallest infinite. But is this second infinity the infinity of the decimals? Can we stretch the decimals to look like this? The big problem is the whole idea of stretchings because sequences had a natural stretch as start. But the natural form of the decimals is the continuum, as a point sets of intervals. It doesn't mean they couldn't be stretched, once they are ordered in this way always forward, that is well ordered. And this leads to the second application. Because such totally hypothetical basic stretch though gives absolutely no clue where such stretch of the decimals could land in the world of stretchings, it at least shows that it is somewhere there! Cantor believed both that such stretches of the continuum are possible and also that the minimal of these would be exactly the sequence stretchings. That is, the infinity of the continuum is the next after the sequence. This became known as the Continuum Hypothesis and turned out to be the first and only concrete undecidable problem. A gaping hole in Set Theory, unless it turns out by a wider theory that our present sets have to be like this.
But our subject is the first assumption of Cantor that the continuum is stretchable at all!
He thought that all sets have a well ordering, that is where all beginnings have a next element.
He also knew that this is a crucial step for his whole infinity comparison theory.
Indeed, though the equivalence defines equality of infinites, the smaller or bigger is still a problem. A precise definition of an S set being bigger than T should not just show that they are not equivalent but also show that S is trivially "at least as big" as T.
This means that either T is a subset of S or that at least it is equivalent to a subset of S.
So T can be injected into S by an equivalence.
But saying this is not enough! We have to prove two missing points.
First, that this is not a crazy definition, that is we can not have both S being bigger than T and also T being bigger than S. This boils down to show that if both injections were possible then there could have been no proof for non equivalence. Assuming consistency then it's enough to show that if injections are possible in both direction then S and T are equivalent too.
This is the so called "Equivalence Theorem".
The second part is that this definition of bigger and smaller should be applicable to all sets.
This means to prove that for any two sets one is definitely equivalent to a subset of the other.
That is where the stretchings and them being beginnings of each other can help.
All we need is that every set can be stretched. Cantor regarded this as intuitively trivial.
As I said the forward stretchings of Galileo are a bit different because we start from an already given stretch. To make one arbitrary stretch from a bag of objects, we should pick out elements again and again until the bag is empty. Zermelo realized that this is actually provable.
A crucial point is that this picking again and again relies on a partial concept of time.
Partial only, because physical time can only flow in the decimals and so for bigger sets the use of time is unacceptable for a physicist. Probably there is a deeper concept of time beyond the real numbers but to appeal to such is even worse. So if all this presently "nonsense" rambling is avoidable then hallelujah we found something solid.
Nowadays we only go in these directions and the paradoxical is not enough to lament about.
I myself didn't mention the most paradoxical thing about the continuum. It is stretchable in a totally different sense, in the natural meaning of the continuum as interval by any projection:



It's interesting that in the new well ordering stretches, a basic stretch is the whole problem.
So we just claim that there has to be a smallest well ordering of the continuum and this minimal is thus pretty hollow. Here at intervals, there is a maximal, the full line.

The points as objects without elements itself were abandoned in axiomatic Set Theory later.
What the Greeks really thought about the points is pretty foggy because the philosophers and the mathematicians did differ. Newton's skeptical and sobering approach existed already then.
But I myself want to stick to proofs now because we can say lot with these too.
Zermelo's Well Ordering Theorem corresponds to Rice's Theorem! It showed that the concept of Sets as spatial collections is enough for the big discovery of Cantor, the equivalences.
We can exorcize time! It doesn't mean that time may not crawl back later and turn out to be unavoidable. But for the comparability of sets, time is not necessary.
The big catch was that Zermelo also unearthed a still lingering unconsciousness in Set Theory.
A hidden axiom was discovered that we used before but weren't aware of.
This Axiom Of Choice is luckily also a perfectly spatial concept, picking out elements from sets simultaneously. But unlike the earlier collections by properties, here we do the pickings usually randomly! You may say, that's nice, we just got rid of one messy concept, time and the price is bringing in an other one, randomness.
But don't worry because nobody will tell you that here at the Axiom Of Choice randomness is involved. The taboos and the stupid epigones of education will never mention this because they do not even see that. Simply because this randomness is not relevant and so was already omitted as mentionable by the discoverers. The parrots don't think for themselves.
The picking of elements from a set of sets could be regarded as a mere set and called as a sample from the set of sets. But the problem is that these sets may have common elements and then picking same from different sets would collapse into a sample that only contains the non repeated pickings. As a drastic and simplest case of leaving free choices means that we must have at least two possible choice in every set. If all these are the same in every set then of course we need some distinguishable elements too. So we may have the naturals and add to each, two fix elements like yes/no or head/tail or **0**/**1** :

{ { 1 , **0** , **1** } , { 2 , **0** , **1** } , { 3 , **0** , **1** } , . . . This can be abbreviated as a sequence:

({ **0** , **1** } , { **0** , **1** } , { **0** , **1** } , . . . **)**

Observe that this as set would collapse into the single same element but in the sequence we have the hidden index elements.
Now a sample using only **0** or **1** pickings would also collapse and so we regard here too the full sets as variable. So the sample is not a set rather an f function:
f { 1 , **0** , **1** } = **0** or **1** , f { 2 , **0** , **1** } = **0** or **1** . . .

The Axiom Of Choice is not required here to show that such choice function exists because we can find a simple example like picking always the **0**.
A more interesting case where still the axiom is not needed is if we have any set of sets all containing only naturals. Indeed, we can simply say let an f be f (S) = min S.
So as we see we only required an existence and that's why the randomness of such choice functions when it is random, can remain in the shadows.
But most amazingly, even the Axiom Of Choice can be left out of the Well Ordering Theorem!
This sounds crazy because it depends on it and yet it's true because it can be applied as the last step to get what we want. The really last step to get what we really want, the comparison of infinites is also just an application if we look wider.
So all the richness of the Well Ordering Theorem and its applicabilities are a separate line, totally independent of the Axiom Of Choice and equivalences.
The crucial realization to see this line is that functions in general quite apart from how they are created like by rules or randomly as choices have their innate dynamics of growing. Plus this intuitive concept of growth has its dynamics of how properties inherit through them.
So the proper title of this whole article should be: Growth Induction.
Growth and its induction are timely concepts offering their almost self evident facts.
And yet these are replaceable by spatial set collections and so can be proven exactly.

Growth Induction

To start didactically, we should start from the second word "induction". At natural numbers it is easy to see that the primitive one by one $P(n) \rightarrow P(n+1)$ induction is enough to derive the so called "complete" induction that uses all $P(1)$ and $P(2)$ and **. . .** and $P(n)$ cases to imply $P(n+1)$. We can avoid these dots because we can use the $<$ relation, so the cases up to $P(n)$ can be expressed by a single formula.
The basic operations also use one by one induction of their cases.
These two, the defining of objects inductively and implying claims inductively can be generalized to all infinites. At the crucial step of going beyond the sequence, the new element is definitely not having a previous element. So we need a new complete induction as basic form.
Getting the new cases from the earlier ones as a set.
As it turns out this is only our vision because every function can be regarded in this manner.
So not as primitive element giver from a last element, rather from the already collected ones.
We simply must regard the potential elemental widening of any set by any f function.
The failing of f and the accumulation of the collection then becomes definable:
Let f be an arbitrary set theoretical function! Thus for any S set f(S) is either not defined or is also a set. The combining or union of two sets is $S \cup T = \{ e ; e \in S \text{ or } e \in T \}$.
So we collected all elements that are in S or T.
$\{ T \}$ denotes the set having only T as element. Using this, we can express the adding of a single new element to an S set as: $S \cup \{ T \} = S + T$.
This element adding is crucial for us because we define: The f widening of S as:
$S^f = S + f(S)$ if f is defined on S or as $S^f = S$ if f is not defined on S.
$S^f = S$ could happen for two reasons. Either f is not defined on S or f is defined on S, but: $S + f(S) = S$ which means that $f(S)$ is an element of f already, $f(S) \in S$.
If $f(S) \notin S$ that is $S^f \neq S$ then we call S^f a proper f widening.
This terminology is following the one for subsets:
$S \subseteq T$ is defined as $e \in S \rightarrow e \in T$, which allows the reverse, that is $e \in T \rightarrow e \in S$ and then S and T have the same element and are thus called equal.
If we know that $S \subseteq T$ plus that T has at least one element not in S then we write $S \subset T$ and say that S is a proper subset of T. A more visual way to say this is that S is narrower than T or T is wider than S.
The subset relation has a strange feature. Namely:
If S has no element at all, then $e \in S \rightarrow e \in T$ is true for all e because the condition is already false and thus we regard the implication true.
So then all elementless S sets should be subsets of any T. Then, all points should be subsets of any T set, which is absurd. Axiomatic Set Theory avoids this contradiction, because it only accepts sets built up from elements, that themselves, have elements. But this is contradictory again, if we don't allow at least one elementless set. And indeed, that's what Set Theory does.
Everything is built from the single elementless set, the empty set $\varnothing$.
This seemingly stupid formalism has its advantages, but when we think about sets, we must go intuitively first and then use the formalism explaining its true meaning. That's how I will go.
For example, our f widenings should start from a chosen, also single elemental $\{s\}$ set and so our "stages" would be:

$$S_1 = \{s\} \quad , \quad S_2 = S_1{}^f = \{s\}^f = \{s\} + f(\{s\}) = \{s\} + f\{s\} = \{s, f\{s\}\} \quad ,$$

$$S_3 = S_2{}^f = \{s, f\{s\}, f\{s, f\{s\}\}\} \quad ,$$

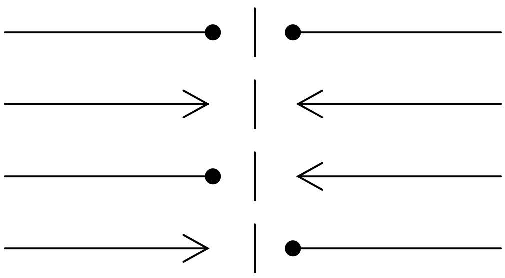
$$S_4 = S_3{}^f = \{ s, f\{s\}, f\{s, f\{s\}, \{s, f\{s\}, f\{s, f\{s\}\}\}\} \quad , \quad \textbf{. . .} \text{ and so on.}$$

If we encounter an $S^f = S$ situation, we regard it as a "termination" of our earlier proper widenings. But such $S^f = S$ in itself shouldn't be called a termination, because such S sets can be plenty, though never reached by the proper widenings, starting from $S_1 = \{s\}$.

If S_n is our termination stage, then $S_1 \subset S_2 \subset \ldots \subset S_n = S_{n+1} = S_{n+2} = \ldots$
If no termination happens after finite widenings then, $S_1 \subset S_2 \subset \ldots$
This situation shows first time the explained heuristic usage of f as set widenings.
We must now form a new $S_\omega = S_1 \cup S_2 \cup \ldots$ stage that collects all elements present in the earlier stages. So actually this S_ω is not just a union, but a limit of the earliers.
From this S_ω limit stage then, we can again start f widenings.
This is the grand vision of "growth". Every f function grows by itself if we let it start from an $S_1 = \{s\}$. The growth never stops, but the proper widenings do when a first $S^f = S$ happens.
And so we call this as the termination.
As I mentioned, $S^f = S$ can be for many S sets that won't become our termination.
So the G growth is defined intuitively as the set of all stages up to termination, but the termination itself is a big unknown. Luckily, it is quite easy to obtain from G. It should be the combined stage of all stages in G, and such combining of a set of sets is easy to define as:
$\bigcup G = \{ e \; ; \; \exists S \text{ that } e \in S \in G \}$
This indeed collects all e elements for which exists an S that $e \in S$ and $s \in G$.
So, $\bigcup G$ is the widest stage in G, the termination stage.
To describe G itself, we could just formulate the three kinds of elements collected in the stages.
The start, the f widenings, and the union widenings:

G^1 : $\{s\} \in G$
G^2 : $S \in G \rightarrow S^f \in G$
G^3 : B is a beginning of $G \rightarrow \bigcup B \in G$

The obvious problem is that we didn't define precisely what such B beginning should be.
We might say, it is simply the stages in G that are narrower than a T stage, so:
$B = G(T) = \{S \; ; \; S \in G \text{ and } S \subset T\}$
The problem with this is that though it is correct, it is useless in G^3.
Indeed, this rule exactly wants to say that this next T stage exists in G, so using T in the condition would claim nothing.
Can the beginnings be defined without the next stage? Yes they can!
The simple but heuristic idea was discovered by Dedekind. His aim was not these widenings, rather the continuous line. And continuity itself can be defined by his heuristic method.
The points of course seem like elementless objects, and thus non acceptable. But we can also regard them as, for example, infinite decimals. These inner identifications of the points are irrelevant for Dedekind's vision. The "point" is merely that the points are ordered by the left right relation, just as our stages are ordered by the $\subset$ relation, though we didn't explicitly claim such ordering yet, by rules.
The grand idea is the concept of "cut". To simply separate a full ordered set in two, the left or B beginning and the right or E end section. The definition of this is simple:
Every element of B must be left from every element of E.
Then amazingly, we can only have four options for a cut:



This picture speaks for itself! The separating line is not in the set, only the lines.
The first case means that the beginning and the end both have extreme members. That is, B has a maximal and E has a minimal. This is also called as a discrete cut.
The second scenario means neither having such elements, so both approaching each other, so both B and E being limit cuts. This is also called as having a hole.
The third means B having maximal but E being limit. Here, the maximal element of B is the one that is approached by E, so there is no hole.
The fourth means B being limit but E having minimal.
The natural numbers ordered increasingly, will only have the first kind of discrete cuts.
The rationals or fractions can have the last three. Indeed, the cut can be at any irrational, making a hole, or at a rational in B or in E.
Dedekind's realization was that continuity simply means to have only the third or fourth kind of cuts, that is, never having discrete gap or a hole.
Our forward going widenings in G can only have the first or the last. So we must always have a next element after a beginning.
Strangely, this whole detour was irrelevant!
We can simplify our third rule as: $H \subseteq G$ and $H \neq \emptyset \rightarrow \bigcup H \in G$.
Indeed, every H subset of G goes up to a B beginning and then $\bigcup H = \bigcup B$.
Simply because the missing stages from H, that is the ones present in B but not in H, will have a wider stage in H and so their elements will be collected too. This more universal claim visibly contains the $H = G$ possibility in our rule. An indeed, we want G to go till termination, so this $\bigcup G \in G$ claim should be true too. A more hidden possibility is that H goes up to a beginning that has a maximal stage in it, which then has to be $\bigcup H$. Then, since $\bigcup H \in H$, thus, obviously $\bigcup H \in G$ too, so we just claimed something trivial.
Our three rules for G only tell what must be collected, but do not exclude any junk.
So in truth, a G that obeys our rules should be called a growth containing set.
The solution to get our intuitively imagined junkless growth lies in the concept of common path or intersection, the parallel concept of union: $S \cap T = \{ e \, ; \, e \in S \text{ and } e \in T \}$.
So we collect those elements that are elements in both. For a whole set of sets, the definition is a bit trickier: $\bigcap \mathcal{G} = \{ S \, ; \, \forall S \, (G \in \mathcal{G} \rightarrow S \in G) \}$
So this collects those S elements that are elements of every element of $\mathcal{G}$.
I used these letters because if $\mathcal{G}$ is the set of all G-s that satisfy our rules, then indeed, this $\bigcap \mathcal{G}$ is exactly those S stages that are present in every G.
So we obtained the junkless growth as $\bigcap \mathcal{G}$.
We made a hidden assumption that this $\bigcap \mathcal{G}$ is itself a G. But observe that our rules trivially inherit to intersections, so our assumption was correct.
Thus, we succeeded in eliminating time from our intuitive vision and obtained the growth as a purely spatial set collection. The problem is that this solution is so abstract that we don't know much about this $\bigcap \mathcal{G}$. It must be trivially contained in every concrete G but we don't have good concrete examples for G-s. A simplest concrete example could be to collect all sets.
Then the rules are obeyed trivially, but we learnt nothing about $\bigcap \mathcal{G}$. In fact, at the start of Axiomatic Set Theory, it turned out that this collection of all sets as a set, is contradictory.
A more reasonable G example would be to collect "only" all those sets that have the s starting element as element. It's easy to see again that our rules stay for this G. But here we learnt something from this G, namely that in $\bigcap \mathcal{G}$ too, all stages must have s as element.
Indeed, $\bigcap \mathcal{G} \subseteq G$ so all stages in $\bigcap \mathcal{G}$ must be in G, that is be "G kind".
This twisted way to say it, is useful in general too. Indeed, any P (S) stage property can be regarded as a collection of stages, that is as $G = \{S \, ; \, P(S)\}$.
And if this is indeed a G, then $\bigcap \mathcal{G} \subseteq G$ implies by definition of G that all $\bigcap \mathcal{G}$ stages must obey P. Of course, $\{S \, ; \, P(S)\}$ to be a G, simply means that :

P^1 : $P(\{s\})$

P^2 : $P(S) \rightarrow P(S^f)$

P^3 : $P(S)$ for all $S \in H \rightarrow P(\bigcup H)$

The second rule seems like a simple induction but actually corresponds to complete, because the stages are already widening. Every new f stage is an increment.
The third rule allows set stages that are not extensions by new f values yet.
As I mentioned, in Number Theory, complete induction follows at once from simple.
Here, these P properties that can be derived are useless and correspond merely to the mentioned trivial G-s. The good news is that we do have a third kind of induction beyond complete and this is our whole subject and title too. This more general "beginning" induction in $\bigcap \mathcal{G}$ regards not P properties about the S stages, rather about whole set of stages. So this grand idea could be seen already among the natural numbers by using not
P (1) and P (2) and . . . and P (n) , rather P {1 , 2 , ... , n} .
But in Number Theory, we exactly want to avoid such use of sets.
Here in Set Theory, this step is the natural. So properties of the full beginnings must widen just as the beginnings as objects themselves widen. The end conclusion is not the weak claim that all stages in $\bigcap \mathcal{G}$ obey some P, rather that $\bigcap \mathcal{G}$ itself obeys a P. So if:

P^1 : $P(\{\{s\}\})$

P^2 : $P(\bigcap \mathcal{G}[S]) \rightarrow P(\bigcap \mathcal{G}[S] + S^f)$

P^3 : $P(\bigcap \mathcal{G}[S])$ for all $S \in H \rightarrow P(\bigcap \mathcal{G}[\bigcup H])$

Then $P(\bigcap \mathcal{G})$.
The used $\bigcap \mathcal{G}[S]$ denotes $\{ T ; T \in \bigcap \mathcal{G}$ and $T \subseteq S \}$ that is, the beginning up to S including S too.
Here we can use these closed beginnings defined by the S maximal element, because now the S existence is not the claim, only the inheritance of P.
We might think that just as in Number Theory, the complete induction was an easy consequence of the simple, here too, this beginning induction could follow from the complete by regarding the S stages that obey our rules. But this is false!
The problem is not that $\bigcap \mathcal{G}[S]$ is a set, and not a single stage, rather that this set is depending on $\bigcap \mathcal{G}$. So we can not regard these stages as a G because $\bigcap \mathcal{G}$ is also depending on the G-s. We would have a circulum virtuoso. The only way out is a long road.
We must avoid $\mathcal{G}$ and $\bigcap \mathcal{G}$ and regard the $\bigcap \mathcal{G}[S]$ closed beginnings as C sets obeying some $C^1 \ldots C^n$ rules. Then again, we can form the $\mathcal{C}$ set of all C-s obeying these rules.
But observe at once that now forming $\bigcap \mathcal{C}$ would be meaningless to avoid junk, because it just gives the smallest C which is $\{\{s\}\}$. So our heuristic junk avoidance is lost!
This has to be achieved by our $C^1 \ldots C^n$ rules.
But if we succeed in this, then $\bigcup \mathcal{C}$ should become $\bigcap \mathcal{G}$.
Using union instead of intersection, unfortunately also means that a trivial inheritance of the rules to $\bigcup \mathcal{C}$ will not be true. This, that is $\bigcup \mathcal{C}$ being a C has to be proven.
This will also show that all C are beginnings of $\bigcup \mathcal{C}$ and of course it also means that $\bigcup \mathcal{C}$ is actually the widest C. This will easily show that $\bigcup \mathcal{C}$ is a G too, so $\bigcup \mathcal{C} \supseteq \bigcap \mathcal{G}$.
In addition, we'll prove that every S stage in $\bigcup \mathcal{C}$ is present in every G, so $\bigcup \mathcal{C} \subseteq \bigcap \mathcal{G}$.
Thus, $\bigcup \mathcal{C} = \bigcap \mathcal{G}$ and the C-s are indeed the closed beginnings of $\bigcap \mathcal{G}$.
Then finally, we can regard the P inheritance rules for C-s:

P^1 : $P(\{\{s\}\})$

P^2 : $P(C) \rightarrow P(C + (\bigcup C)^f)$

P^3 : $P(C)$ for all $C \in H \rightarrow P(\bigcup H + \bigcup\bigcup H)$

We can again regard this P as a set of C-s that obeys this.
Since the P rules will follow from $C^1 \ldots C^n$, thus $\mathcal{C}$ is a subset of this P as set.
So all C-s obey P including $\bigcup \mathcal{C}$ that is $\bigcap \mathcal{G}$.
The real "ad-hoc" step of course is to choose $C^1 \ldots C^n$. And yet, they can only be perfect.
If they are too strong, then $\bigcup \mathcal{C}$ is too small and wouldn't cover $\bigcap \mathcal{G}$.
If they are too weak, then $\bigcup \mathcal{C}$ is too big and we would have stages not in $\bigcap \mathcal{G}$.
So the finding of $C^1 \ldots C^n$ is quite tricky. They have to be simple so their inheritance is easy to prove, but they have to be strong to prove this inheritance, plus the second claim of all stages appearing in G-s. To help to find $C^1 \ldots C^n$ is to reveal two main claims that are in between $C^1 \ldots C^n$ and the mentioned two results about $\bigcup \mathcal{C}$.
So these will be proven from $C^1 \ldots C^n$ and will easily yield the two claims about $\bigcup \mathcal{C}$.
We might say, why don't we just take then these in between claims to be $C^1 \ldots C^n$.
Again, because then they are too strong, so the inheritance would be hard to prove.
Another circulum virtuoso would be to use our new induction principle to prove these claims.
Indeed, this new induction is exactly the consequence of our following proofs.
So there is no way out! First we reveal our two claims 1.) , 2.).
These help to find $C^1 \ldots C^n$. Then we prove 1.) and 2.) and then we prove the two claims about $\bigcup \mathcal{C}$ and finally our induction is an easy consequence.

1.) $K \subseteq C$ and $K \neq \varnothing \rightarrow \bigcap K \in K$

2.) $T \in C \rightarrow T = \{s\}$ or

$T = S^f$ with $S \subset T$ or

$T = \bigcup H$ with $S \in H \rightarrow S \subset T$

We should explain their true meanings first.

1.) says that for any K set of stages in a C, there is always a narrowest member in K.
Indeed, if there is such, it has to be $\bigcap K$. This claim seems like an analogue of the union widenings and it is even true in how we'll prove it. Remember that the generalization of the union widening went by regarding the B beginning for any H. Here, an E end will show the claim for K. But also observe a drastic difference. $\bigcup H$ was claimed to be in G and similarly will be claimed to be in C, but $\bigcap K$ must be in K itself.
2.) says that every stage in C must come from being the start or f widening or union widening.
For a second we might think that we hit the jackpot and this claim alone could avoid any junk.
The added nuance of claiming the f widening to be proper does avoid S junk stages, where we simply define f so that $S^f = S$. But surprisingly, we can have junks that are all proper f widenings. To see this, visualize a sequence of narrowing stages that are all f widenings in reverse. Then these come from "nowhere". They can finish by a termination or widen infinitely too. Such backwards narrowing or double infinite sequences could then be added to any C and still obey 2.). They could even contain our real intended C.

Luckily, observe that 1.) forbids such narrowings and so we may feel that 1) plus 2.) is the perfect junk avoidance. We can even improvise a "proof" for this:
Suppose 1.) and 2.) holds and we had a J set of junks.
By 1.) $\bigcap J$ is the narrowest junk, so it has no proper subset that is junk. But by 2.) $\bigcap J$ must be also the three options. The second and third are using narrower S, that is no junk.
But then $\bigcap J$ couldn't be junk either. This was not a correct proof. The correct one will use the same argument, but showing that $\bigcup \mathcal{C}$ has no extra stages outside $\bigcap \mathcal{G}$.
Now we choose $C^1 \ldots C^n$. The G rules are all okay, except that the relentless f widening in G^2 is not true for the widest element of C, that is $\bigcup C$:

$$S \in C \text{ and } S \neq \bigcup C \rightarrow S^f \in C$$

Now we strengthen this weakened version by injecting the essence of 2.) into it.
That is, we claim that for the non maximal stages, the widening is proper:

$$S \in C \text{ and } S \neq \bigcup C \rightarrow S^f \neq S \text{ and } S^f \in C$$

We must also claim the essence of 1.) . As we revealed, it is that the ends have first stage.
This follows if the f widenings are the next stage after a stage without in-between T stage.
And such $S \subset T \subset S^f$ is impossible because S^f has only one extra element beyond S.
Of course, we were thinking in chains again, so it's time to regard this as our final fourth rule.

C^1: $\{s\} \in C$

C^2: $S \in C \text{ and } S \neq \bigcup C \rightarrow S^f \neq S \text{ and } S^f \in C$

C^3: $H \subseteq C \text{ and } H \neq \varnothing \rightarrow \bigcup H \in C$

C^4: $S, T \in C \rightarrow S = T \text{ or } S \subset T \text{ or } T \subset S$

Now we prove 1.) and 2.) .
To show 1.) we define $B = \{s \;;\; T \in K \rightarrow S \subset T\}$
It's a beginning, though might be empty too. $E = C - B$ is the end that might be the full C.
In these extreme cases $\bigcap K = \bigcap C = \{s\}$ by C^4.
If B is open, that is $\bigcup B \notin B$ then $\bigcup B \in E$ is the minimal stage in E and in K too.
If B is closed, that is $\bigcup B \in B$ then by C^4 and C^2 $(\bigcup B)^f$ is the minimal stage in E and in K too.
To prove 2.) we again regard beginnings in C but now for every $T \in C$ we define:
$C(T) = \{S \;;\; S \in C \text{ and } S \subset T\}$ This could again be open, if T was $\bigcup C(T)$ or closed if $\bigcup C(T) \in C(T)$ and then $T = (\bigcup C(T))^f \supset \bigcup C(T)$.

So we arrived to proving our big claims.
The first is that $\bigcup \mathcal{C}$ is a C, that is $C^1 \ldots C^4$ inherit to $\bigcup \mathcal{C}$.
We show that the rules inherit to $\bigcup \mathcal{C} + \bigcup\bigcup\mathcal{C}$ for any $\mathcal{C}$ set of C-s in general.
We had to add $\bigcup\bigcup \mathcal{C}$ because for arbitrary $\mathcal{C}$ the union might not be closed.
For $\mathcal{C}$ being all C-s this is irrelevant. Indeed, $\bigcup\bigcup\mathcal{C}$ has to be already in $\bigcup \mathcal{C}$ because otherwise we could get a wider C by adding $\bigcup\bigcup\mathcal{C}$ to $\bigcup\mathcal{C}$.

This more general inheritance proof relies on an even more general fact about C-s, namely that for any two C_1 , C_2 one is a beginning of the other.

Surprisingly, to prove this, we'll use a simpler version of our general inheritance, namely, when we know that all C members of the $\mathcal{C}$ are beginnings of a C*.

The inheritance here is quite simple.

Either all stages in C* appear in some C in $\mathcal{C}$ and then trivially $\bigcup\mathcal{C} = C^*$.

Or if there are some stages in C* not appearing in any C from $\mathcal{C}$ and the narrowest such is S, then $\bigcup\mathcal{C} = C^*(S)$.

So now back to why for any C_1, C_2 one is beginning of the other, we can regard the common beginnings of C_1 and C_2 combined. By our above simpler inheritance, this combined set is either the full or a beginning in both C_1 and C_2.

It can't be the full in both, because then $C_1 = C_2$ were. It can't be the proper beginning in both either, because then we could make a wider common beginning.

So this combined set is the full in one and a beginning in the other, which exactly means that one is beginning of the other.

Now we can prove the general inheritance for any $\mathcal{C}$.

C^4 is trivial because for any two S , T if they came from C_S and C_T elements of $\mathcal{C}$ then one is beginning of the other, so the wider one contains both S , T already, where we had $S \subset T$ or $T \subset S$.

To prove the inheritances of C^1, C^2, C^3 it is enough to show that for any $T \in \bigcup\mathcal{C}$, if C_T is any element of $\mathcal{C}$ having T as element, then $C_T(T) = \bigcup\mathcal{C}(T)$.

So the beginnings before T are the same in $\bigcup\mathcal{C}$ and C_T. Then C^1, C^2 remain trivially and C^3 could only fail if we combine stages for which there are wider ones in any C in $\mathcal{C}$.

But adding $\bigcup\bigcup\mathcal{C}$ solves this.

$C_T(T) \subseteq \bigcup\mathcal{C}(T)$ is trivial so we only have to show the reverse, that is:

$S \in \bigcup\mathcal{C}(T) \;\rightarrow\; S \in C_T(T)$

$S \in \bigcup\mathcal{C}(T)$ just means that S came from a $C_S \in \mathcal{C}$ and that $S \subset T$.

But we know that all C-s are beginnings of one another.

If C_S is beginning of C_T then $C_S(S)$ is a beginning of C_T too.

S is the next stage after this $C_S(S)$ in C_T too and since $S \subset T$ thus, $C_T(T)$ is a later beginning, so $S \in C_T(T)$.

If C_T is beginning of C_S then $C_T(T)$ is beginning of C_S too. But all stages outside $C_T(T)$ in C_S are the T or wider ones and $S \subset T$ and $S \in C_S$ thus S must be in $C_T(T)$.

Now we return to $\mathcal{C}$ being the full set of closed beginnings and thus, $\bigcup\mathcal{C}$ indeed being the widest C satisfying 1.) , 2.) . As we showed, $\bigcup\mathcal{C}$ has to be terminating and so the C^2 rule can be unrelaxed so G^2 will stand. Thus we obtained an existing G as $\bigcup\mathcal{C}$.

But to show that it is $\bigcap\mathcal{G}$, we must prove that all stages in $\bigcup\mathcal{C}$ must be in all G.

So let J be the set of all stages in $\bigcup\mathcal{C}$ that are not in a chosen G.

$\bigcap J$ is the narrowest element of J, thus a stage in $\bigcup\mathcal{C}$ that is not in G but has no such proper subset. By 2.) any T stage in $\bigcup\mathcal{C}$ is $\{s\}$ or S^f or $\bigcup H$ with $S \subset T$ and $S \in H \rightarrow S \subset T$.

$\bigcap J$ can not be $\{s\}$ because this is in any G by G^1 .

$\bigcap J$ can not be S^f with $S \subset T$ because then S is in G and so by G^2 , S^f would be there too. Finally:

$\bigcap J$ can not be $\bigcup H$ with $S \in H \rightarrow S \subset T$ either, because here G^3 makes this to be in G.

The proof of our induction principle is easy now.

Finally, we arrived at the most famous applications of our whole system to show that all sets can become a growth.
And any two sets can be compared by equivalence of one into the other.
The first is the Well Ordering Theorem using the Axiom Of choice.
Namely, we use a c choice function on all subsets of our A arbitrary set.
$c(A - S) = f(S)$ will be our widening function and $f(A) = c(\varnothing)$ is our starting s.
This f is defined on all proper subsets and the values are always outside.
So a termination can only happen at the only non defined full subset $S = A$.
So indeed the growth will have A as widest stage.

For two sets the existence of partial equivalence is a P property of subsets and it will inherit by our beginning induction to a full set in one of them. Thus for any two sets one is also equivalent with a subset of the other.