The First Stage Of Meta Mathematics

A Very Personal Introduction

This is a very ambitious article because this title hides many meanings.
The most apparent is that I claim a first stage has passed. That something is over and something new is brewing. I even tell right now that the new second stage will be Randomness penetrating Set Theory but I haven't the faintest clue how it will happen.
The fact that there is no article targeting what I do now, that is to explain simply yet fully the first stage, is a tragedy! It is a reflection of our time. The stupidity of the internet and Wikipedia's democratized knowledge "spreading" in effect over formalizing and over complicating all simple truths, is the most concrete background. Understanding is not the goal anymore in this world and not just for the teachers but for the pupils themselves.
And yet the "Didactical Eye" is not dead completely. I can name three people who tried to do what I am doing now but it seems that beside a didactical eye one needs the didactical arse too. Smullyan, Boolos and Odifreddi repeatedly expressed how simple things could be but then fell short of systematically bring down the results to the simplicity with which they could and should be explained. Plus the fact that they didn't reflect on the universal Formalism itself that paralyses not only education but the whole modern spiritual existence is a sign that they were afraid to face the full reality. They were committed to the wrong side.
I have opened my eye in America and only by the help of LSD. Timothy Leary's "establishment" became the most concrete and only important social "abstraction", beyond all political nonsense.
I never "sold out", never became part of any social structure. I was attached to matter and on many levels that I regret, but not becoming a "real" mathematician is one thing I am proud of.
So if you are looking for real professors to explain things then leave right now.
I will not be measured and objective! I will tell it as I see it! This is honesty and it works for spreading not just opinions but knowledge too. Simply because no knowledge of facts can replace the vision of how things seem. How one sees the situations subjectively! The correct vision can easily be detailed to the correct facts but correct facts can hide the big picture completely. This is where the Formalists disagree with me! They claim that proof of facts is all that mathematics should be and the background visions are private business. The situation is actually much worse than this would suggest. Indeed, there is a certain type of character who learns this whole present approach as behavior and when actually gets to form his own visions from the details, becomes even more non expressive than the bad masters were. So the situation is deteriorating in time.
The fact that we have all kinds of Nobel prizes, even economic one in disguise for second grade mathematicians, but not for education, is a perfect sign of how only the new, only the provable facts and only measured by an upper peer group counts as value.
When I was still very young and returned from America to Hungary, my first visit was to Andras Hajnal and told him how I struggled in high school due to the lack of even a modern Set Theory book in Hungarian and I wish to write one and would appreciate his criticism. He smiled and said that those who are really interested can read the foreign ones. My frustration culminated later in a few letters I wrote to Erdős. He replied at once and stood by his Formalism boldly and stupidly.
I left Hungary again and lived in Australia in the last twenty five years. This "lucky country" was not welcoming my didactical eye either. Only the many students I tutored gave me a firm reassurance that I was right and the whole world is crazy.
Now I should go into more of the mentioned hidden claims in the title.
The word Meta is an obvious reference to Kleene's "Introduction To Meta Mathematics". He was the only one who used this word for new math and actually it is the most descriptive and truthful word for it. So I try to revitalize it and maybe it will catch on. His book was a mile stone in trying to grasp the new field and indeed it became a bible. Unfortunately, it was too early. The true simplicity was not yet discovered and was completely missed in his bible. This has a parallel in the Hungarian Rozsa Peter's introductory book "Playing With Infinity". In 1938 it was a masterpiece but to reprint is in 1969 with Csaszar's stupid preface was a disgrace.

The first stage of meta mathematics being a finished period hides more than just the claim that the clarification is now over and should be offered for everybody. The situation is much better potentially! While in classical mathematics the different smart proofs always did and always will contain some ad hoc ideas that simply worked, here in this new meta mathematics the actual proofs completely flow out of the clear vision. So as surprising it is, meta mathematics is simpler than classical. Clarity here is a must because everything is logical.

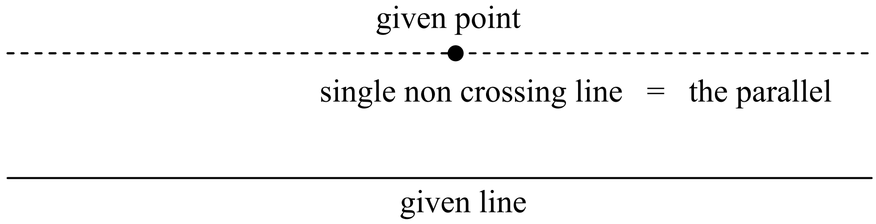
The greatest classical mathematician was Gauss, the greatest meta mathematician was Gödel.

Gauss had many tragic points in his life but two professional ones. The first is that his complex numbers didn't infiltrate physics in his life though he was sure that this would happen. The second one was that his entry into non Euclidian geometry didn't push him to something that should have been so obvious given his mania for exactness of proofs. Namely, the concept of models would have led him straight to meta mathematics. So this second "bad luck" was totally his own limitation in vision. A catch twenty two! His obsession stopped him from fulfilling his very obsession to its limit. A side effect of all this was his insensitivity that almost drove the young Hungarian mathematician Janos Bolyai to suicide. Janos' father was classmate of Gauss and so Janos sent his book to him for evaluation. The short reply was that he (Gauss) could not praise the results because himself achieved them years earlier and left them aside since time is not ripe for them yet. Gauss was not lying and indeed Bolyai was standing with the same blockage at his own results as Gauss was earlier. Janos wrote in his notebook that "From nothing I created a new world!" which is a pretty upbeat reaction to a blockage. This less self critical valuation of a young man than an old and already world famous and acknowledged genius, makes it even more evident that the wise one should have been wise as a human being too. Gauss wasn't.

Anyway, to see this mentioned blockage is vital and so I explain this as the start of our subject.

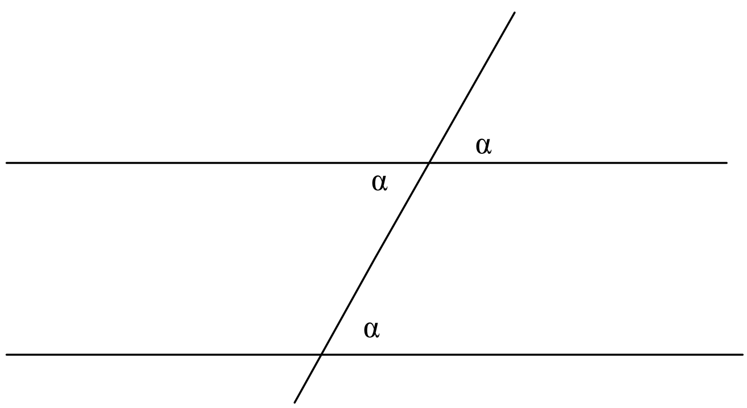
Part One : Derivability

The axiom of parallelity in its modern form says that for any line through any outside point there is only one second line that will not cross the first. This is what we call the parallel.

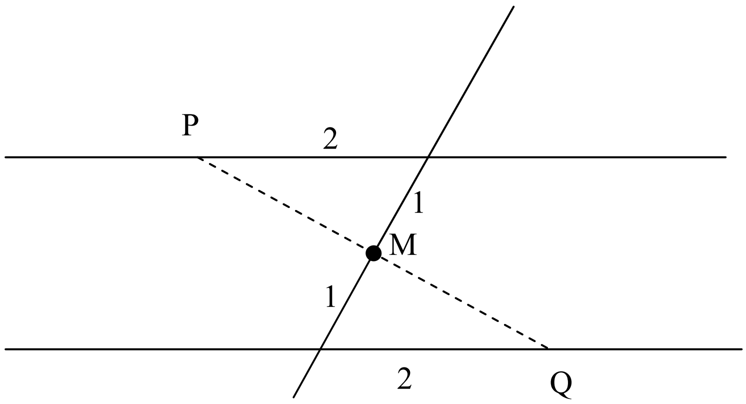


So we had a claim and instantly turned it into a definition.
The beauty of this is that we didn't have to talk about the usual visual features of parallel lines like being fix distance between them or having the same angles when crossed by a third line.
This is interesting because actually the whole point of a parallelity axiom is to prove that these three features, the fix distance, same angel and non crossing are implying each other.
Euclid's original axiom was not as smart as this modern, rather made a claim that if two lines have different angles with a crossing third then they must cross. So, the uniqueness remained hidden until actual derivations brought it out. He was the first reluctant Formalist. He sacrificed didactical clarity for precision of derivation by logic. But lets be specific! The concrete didactical mess is coming from the fact that angles are actually complicated. Two angles being the same means that same distances measured onto the sides and then connected must create again a third same distance between the sides:

Thus the definition of the angle has a similar claim involvement as parallelity has above in the modern definition. Here the assumption is that if we have one such equality of cross distances for two crossing lines, then it will be true for all and this is what we call having equal angles.
The two most obvious claims about non crossings now seem to be easy to prove too.
Namely, that fix distance between two lines or same angle to a third, both imply non crossing.
For two lines with fix distance the claim seems even trivial because at a crossing the distance would have to be zero. But observe that the "fix distance" itself is not so trivial because here distance means minimal from any point to the other line and thus requires right angled connector.
For two lines having the same angle to a third, the trick is to regard the mirrored version of the second angle that is not same positioned rather opposite to the original.

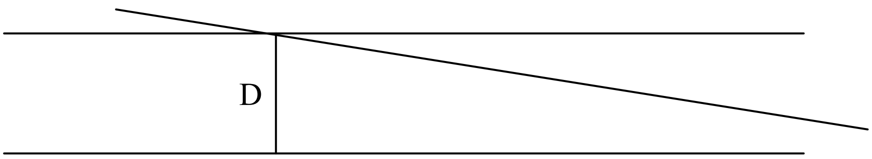


Then, the section of the crossing line between the two lines can be regarded and halved at M. The two half lengths can be used as 1 same distance on one side of the equal angles. Then picking any P point on one line will determine a 2 distance and using this same on the other will determine a Q on the other. And these as second distances on the angles will determine the same long PM and MQ connectors:

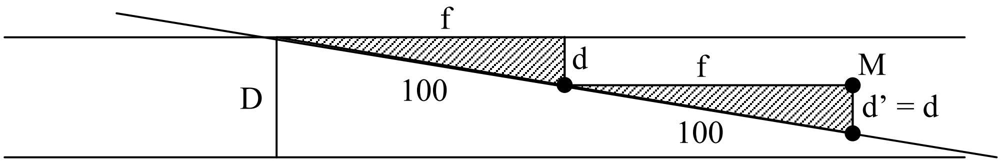


But then actually the two triangles are mirrored as whole and so PM and MQ are not just equal but are aligned. Thus we actually mirrored the arbitrary P point of one line to the other as Q. And thus a crossing P is impossible because it would mean that its mirrored Q is also a crossing and so the two lines would have to be the same, the line through P and Q.

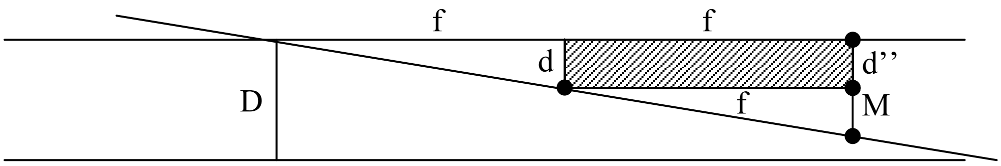
We only showed that non crossings must stand for the other two meanings but if non crossing is unique then actually it means that non crossing implies at once these two in reverse too and as a side result also that these two imply each other too. So as I just said at the start, the three intuitive vision of fix distance, same angle and non crossing are the same.
The root of the problem is that this whole commonness of the three visions seems so obvious that we would expect it to follow from the simpler axioms.
Those simpler axioms say things like: Every two points determine a single line. Two lines can cross only in one point. And so on. If we look closer, actually these are not so obvious already.
For example, we might think that the mentioned two imply each other because:
If two points would determine more than one line then at once we had two lines that cross in two points. Plus in reverse too if two lines were crossing in two points then those two points were determining two lines. But these arguments were faulty! At the first we ignored the possibility that two points would determine no line at all and at the second the possibility that two lines don't cross at all. The first is impossible and so this should be an axiom on its own and the uniqueness as an additional claim. But at the second the situation of non crossing is very possible and leads exactly to parallelity.
The basic vision of the line is actually an extension of the distance as we used it above already too. So the straightness of the line means that it has to be the minimal distance between pints. This is the really hidden agenda that we try to capture by crossings and determinations. And it seemed that we could achieve this but only with the then added extra condition that the non crossing is unique. If this could be derived from the simpler crossing conditions then of course everything would be perfect.
But lets see how we could argue intuitively for a uniqueness, for example at the fix D distance: Lets pick a P point as fix too and thus we keep the D distance here but we slant the line through P a tiny bit. One side goes up the other down and we feel sure that this side, say on the right will eventually cross into our base line:



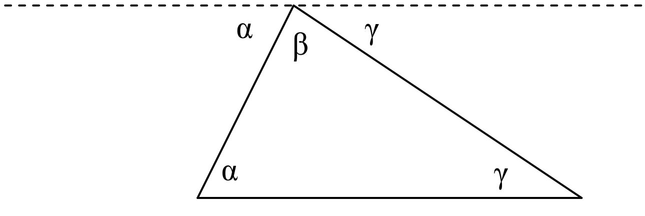
We would even say that this is obvious because it gets closer and closer to the base line.
And here is our first insight. Just because the distance is decreasing it doesn't imply crossing! Indeed, why could it not decrease continually for ever that is never actually getting to zero.
So then we get into the details and say okay lets see how much our slanted line drops after every hundred meter. If it drops a fix d distance then no matter how small this d is, the infinite line is enough to step arbitrary m many hundred meters and thus drop md. But this then will reach the original D distance after exactly $m = \frac{D}{d}$ many drops because $\frac{D}{d} d = D$.
So the crucial point is whether such proportional that is fixed step drop is derivable.
And of course the essence is the second step. Whether we can show there that the drop becomes 2d because then this can be repeated to get 3d and so on.
The first d drop has an angle to the slanted line and we can copy this angle at the next hundred meter by finding the M point that has d' = d distance from the drop and across to the first drop has the same f forward distance as we had at our first drop. So d' is non crossing with d :



Then we connect M with the second forward point to get a d'' distance that is non crossing with d again due to the fix f distance now. The shaded parallelogram also gives that d'' = d :

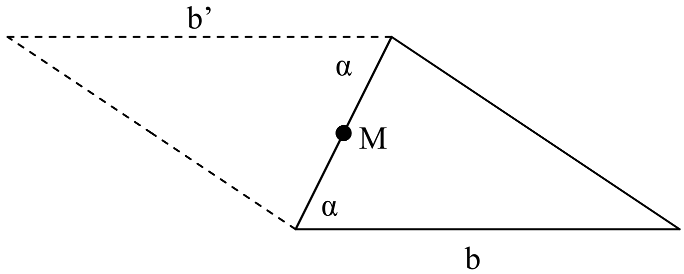


The above parallelogram actually seems like a rectangle but this only follows from our belief that the original right angle direction is kept. Aside from this problem, it indeed seems that the second drop is d' + d'' = 2d. But this is only an actual drop if d' and d'' align which only follows by having a single non crossing line with d through M.
So our whole seemingly so natural proportionality argument requires tacitly the parallelity axiom.
The best example of how smoothly the parallelity axiom can be used hidden, is at the famous theorem that the three angles of any triangle always total 180, that is a flat line.
The usual false "proof" in schools is simply to draw a single line as "illumination".
Namely above through the corner and being parallel with the base line:

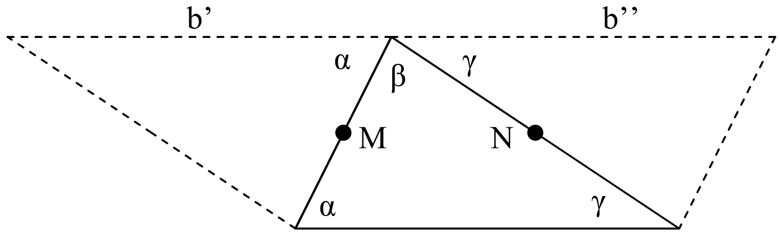


As we see, the corner has already one angle there and the other two were copied there too forming a full 180 together indeed. Of course, the foggy part is why those angles are the same at all.

To do a true copying of just one angle as start, can be established by mirroring the base line to the M middle point of one side:



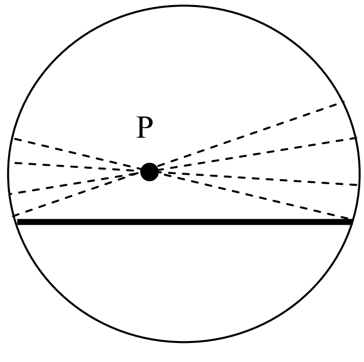
This b' mirrored base line not only will have the same α angle there but if continued as a line, it will have to be a non crossing one with b-s extension. Indeed, the mirroring extends again and a crossing would mean that mirrored points are the same.
Now we can do a similar mirroring for the other γ angle through the N middle point, and so we get a second b'' mirrored base line at the corner too. This b'' again if extended must be a non crossing line with the base line. So, all three angles appear in a seemingly single line again:



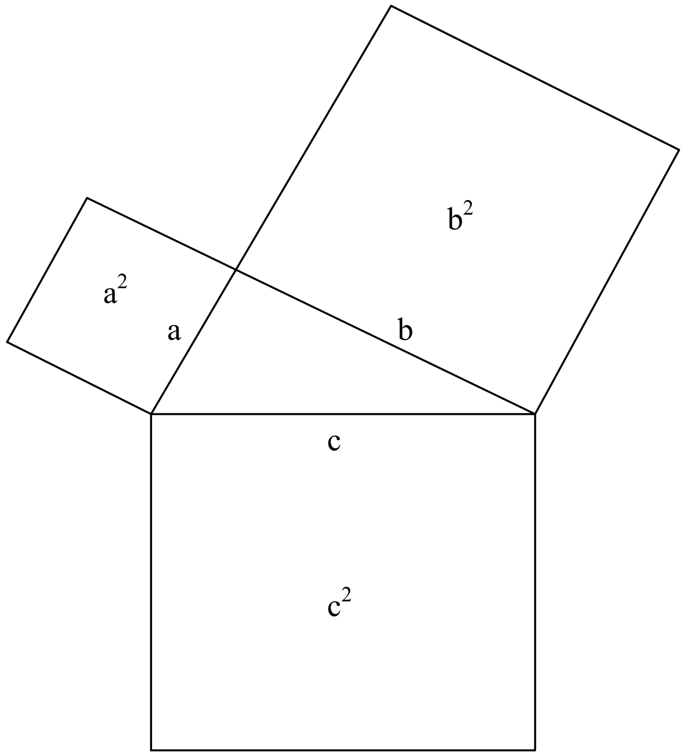
And so finally the "cat's out of the bag". The three angles only form a true line if these two b' , b'' mirrored base lines align. Now if there is only one non crossing line there, then this is true.
Seeing this amazing subtlety how the uniqueness parallelity axiom is used, some were intentionally trying to avoid it, or in fact to assume it to be false, just to see whether a contradiction would come about. If it would happen then of course this would at once give proof that unique parallelity is true and we could avoid the axiom. The surprise became that not only a contradiction wasn't reached but actually the consequences of more non crossing lines were not that absurd. So we enter Bolyai's "new world". The distance can decrease between two lines yet never get to zero.
What neither Bolyai nor Gauss realized is that actually there was even a slight mistake in their whole negative endeavor. Indeed, the true negative of the parallelity axiom says that "there is no unique non crossing line" and so technically should also allow that maybe there is no parallel at all. This of course seems so unintuitive that they simply ignored it as possibility. If they saw what soon I will show about this possibility of no parallels at all, they would have seen the light at once. The real problem was not just that they went by intuition but they regarded it as a physical reality. So when they assumed the more parallels then they truly imagined that this might be the physical reality. This was never expressed, nor what lines should be physically like light rays or something else.
The newer knowledge that physical space-time is indeed non Euclidian might give the impression that then they were even more forerunners as physicists but this is a false line of logic. This Relativistic non Euclidianess is very different from theirs. The Euclidianess is indeed a universal proportionality. A tiny picture of a kitten can not just be enlarged to a cinema screen but up to any galactic sizes. And all internal angles should remain, similarity has no boundaries. This is not true in our universe because it is finite. So an instant projection can not increase without limit. But the universe is not simply finite as a space with boundaries rather it is expanding. This makes the boundary unreachable so a compromise is achieved. We can say finite but we can not get a trivial contradiction from this, like where is the center. Every point could be a center because the edge is

faster. As they say "time is of the essence" and so here too this was the new dimension added to space that made actually space-time non Euclidian.

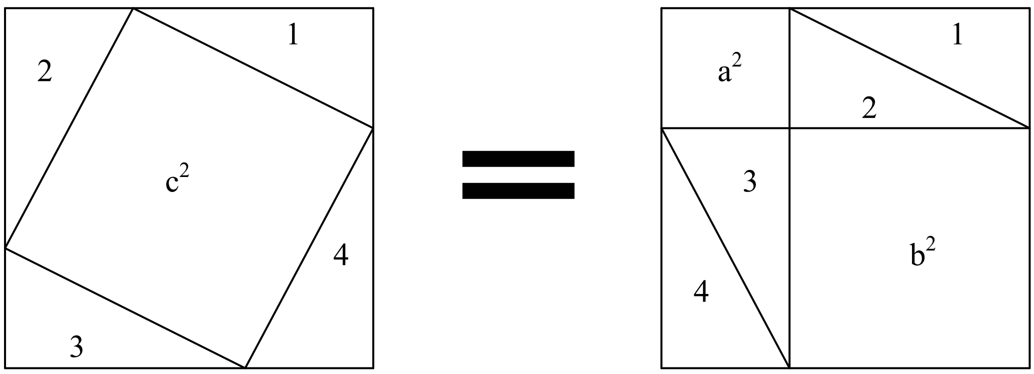
So lets see the mentioned weird assumption that to a line there is no parallel at all, from a whole new perspective! Lets forget real lines and points and just try to create a fantasy world where everything looks different but the axioms of lines are true except of course parallelity.
The amazing thing is that such fantasy world is very real, more real than an infinite plane.
The surface of our globe we live on is a perfect such world if we regard some of the circles as lines! Namely, those so called main circles should be the lines that are the biggest possible. These are the ones that have a center right in the center of our globe. Most famous is the equator and perpendicularly all the time zones. But amazingly there are such main circles for any two points.
For example New York and Moscow will determine one too. First this seems confusing because they are both northern cities and the equator narrowed and moved up could be a connecting circle. But actually this smaller circle can be again increased to equator sized by fixing these cities and turning the circle till its center again goes to the earth's center. This will have a shorter arc going up north and then coming back down, while the other bigger one will go all around the earth down south and back. The shorter route could be regarded as the interval between New York and Moscow while the bigger arc as the weird continuation of the interval to the left and right that miraculously meet each other on the other side of our world. So infinity is tamed in length to finite too. There is even a surprising hidden beauty in this definition of lines and intervals. Namely, these intervals are indeed the shortest possible distances between two points. So, above where we turned the circle into a maximal main one, actually the northern arc decreased. So to fly from New York to Moscow on the shortest flight route is exactly following this arc. It's better to bend towards north and come back. To see this intuitively, first we have to look from above outside our globe so that we stand right above the main circle arc. Fixing New York and Moscow and altering the main circle into other ones, we will go on other circle arcs. Amazingly, even though the full circles become smaller, their arc between New York and Moscow will increase. To see this fact we have to turn the smaller circle into the plane of the main one but keeping New York and Moscow fixed. Then in this plane we simply have two points connected by two circles and we see at once that the bigger circle gives smaller connecting arc. Sadly, a sudden death blow seems to be the following fact. Any two main circles cross each other in two opposite points. For example this New York Moscow main circle will cross the equator in two opposite points. As we know lines can only cross in one point! The solution is simple! We should regard such opposite points of the globe as one point. The north pole and the south pole are together a single point.
In spite of this weirdness. Most plane geometrical features appear on the globe surface very similarly. Triangles for example are three main circle arcs and their angles are the actual angles of the main circle planes seeable as tangents to the main circles. These are a tiny bit bigger than the actual triangle angles inside the earth. This shows that the sum of the angles in a triangle in this surface geometry is more than 180 degree.
This spherical geometry was well known way before non Euclidian geometry started but the whole idea of replacing reality with substitute reality didn't arise. To create a substitute world where there are more parallel lines to one through a single point is a bit more difficult. We have to distort the distances and then they can look like this:



Once the idea was grasped that we can create models in our old Euclidian geometry, the varieties became plenty. These mathematicians didn't even care any more whether real lines are Euclidian or not. If one is real the others are real too as models of realities inside.
I want to emphasize this revolution even more by recalling the old "hindu" method.
This was the avoidance of words and just use drawn pictures in the sand to achieve the proof.
A perfect example is the proof of the Pythagoras Theorem $a^2 + b^2 = c^2$
meaning that the two squares on the smaller sides of a right angle triangle equals to the third one:



We simply start with two copies of a new "ad hoc" even bigger square with side $a + b$.
In one we insert c^2, in the other we insert a^2 and b^2.
"Miraculously", in both case the left over is four copies of the triangle:



This of course indeed proves our claim instantly.

Obviously, if we use modelings and for example we regard circles as lines then such visual method seems impossible. But actually, it is not a true opposite either. A true opposite would be a verbal line of arguments or actually not even that, rather a logical step sequence that formally achieves the claim. This was not possible neither in ancient times nor in the 19^{th} century when these models were discovered. It is possible today though we never actually present the exact

logic. Instead, we pretend that we do. We over formalize and avoid all visuality. The present is an intentional lying and deception. The motivation is not directly to confuse others rather to hide our own confusions. To avoid asking some questions. The modern existence is a continuous avoidance of questions that do not lead to socially rewarded answers. The end result is still suppressing not just our own but other's understanding too.
The modeling itself is merely a very exact translation that shows that there has to be an exact logic but it doesn't show that logic yet. And indeed, once we accept that we are working in a spherical geometry, we can instantly argue as in real geometry but still not be conscious of exactly our steps. There is a very important analogy of this situation with games. The big difference is that in games we establish artificial rules and those are spelled out explicitly at the very start. But the instant translation is amazing here too! We explain the chess rules for example to a child and instead of playing a lot we go out into the park and see a giant chess board there. The child will not only see that the huge pieces may resemble the little ones at home but will be able to play at once. Lifting up the pieces and move them correctly. All this happens instantaneously! Rule transference is a human a priori ability. An other name could be for this "finite isomorphism" principle because finite similar formed structures are recognized. It's an interesting question which is the more important factor, the similarity or that there are rules behind. One thing is for sure! Animals do not possess neither! They live in direct reality. An argument that suggests that isomorphism is the main thing is observing how a child recognizes his father on a picture. The point is not that he recognizes the father rather that he knows it is not the real one. An animal can recognize someone on a film but then it is regarded as real and when it turns out to be fake the recognition is lost again and again. The rule aspect still is a factor on its own too. For example, a rat can learn a maze very fast. If the maze had rules say left, right, left, right leading to success and we use a similar ruled one but with different space lengths then the rat has to relearn it. Rule transference will never come about. But the rules can be learnt gradually and this will give a false impression of seeing the rules. That's the usual lie in animal intelligence experiments.
This half baked analogy of the subconscious logical steps and the spelled out rules of a game will be amazingly the start of the second part of this book, the non derivability. But the real start of this first part is also related to it. Indeed, there was a crucial point in these model creations that can be regarded as the exact start of "Meta Mathematics".
It was Beltrami who first realized that all this fun and game with models implied something trivial about the original quest. Remember that the new world of more parallels were originally chased to get a contradiction that would of course mean a derivation of unique parallels and thus an avoidance of the parallelity axiom of Euclid or any other versions. Beltrami realized that strangely now these models with non unique parallels, actually prove that Euclid was right and so if we want a world with unique parallels then it requires an axiom for that. Here is the argument:
If the unique parallelity would follow from the simpler axioms then also in every model where the other axioms are true, parallelity should be unique too. But our models show that it is not true so the other axioms can not imply unique parallelity. And thus neither any axiom that claims this.
The hidden assumption here was of course that any reality must obey logic.
That's amazing because there was no exact logic at this time yet!
This whole recognition spawned the search for such exact Logic.
As it slowly started to form, a "blast" from the heavens penetrated the process.
Sounds like some mythical creation legend and we are not far from it.
This "blast" was Cantor's "Set Theory" and amazingly the original blast was not relating to the logical line that I pursued above and as a spark attributed to Beltrami. What's more, beside the gradually awakening logical connection of Set Theory, there was still a third much more apparent cultural consequence of it. We have to start with this because it is practically a taboo.
This most turbulent century of our history just behind us is enough to make us dizzy if we try to make a stocktaking of it. My previous words tell what I regard as most important. The emergence of Formalism spreading a hidden materialism. The concrete mechanism of this is the new splintered section of society Media. While the actual actions in political and economic systems already had their silently materialistic influences for thousands of years, this new consumable culture should have been a revolt against it. And it is! But exactly this the twisted situation that it

pours over the individuals with false answers or rather it answers to no questions at all, rather stops the questioning itself in the individuals. So to simplify it, our globe became "Planet TV and Internet" where from now on all generations will be manufactured by programmed awakening of the consciousness. It is a simple statistical fact. Whether the exceptional can escape this and so a revolt against stupidity is successful, is secondary. The anti social intelligence can even be an accepted stereotype. But the simple truth is that everybody has the same intelligence that is socially distorted.
Two World Wars and a new world order of nations are the historical turbulence that brought out and also covered up the emergence of our new Media universe. But now the existence is trivial.
The intricate details of what this homogenized world thought order includes is impossible to stock take yet.
But if this beast is so complex then how do I dare to regard it as a menace and not a bless.
This is a legitimate question an I have to surrender to it. I am a product of this age and I am the last person to advocate restricting a child and not to expose him or her to TV and the Internet.
So, stupidity allows us to become smart and then wisdom is to recognize one's own lies.
The good news is that this whole process was predicted and thus maybe a new turn for wisdom will emerge. This also reveals a common general falsity underlying the new social lie. It could be called the "historical blindness". Though the historical events are continually portrayed in the cultural awakening of the individuals in much more detail than before and thus again suggesting that after all things are better, these visions of the past are always translated into our modern though system. A not quite explicit but definite attitude is that we are much smarter now than before. Combined with the obvious advantages of present life, this is an almost impenetrable delusion. So, I start to believe that my best friends should be not revolting mathematicians rather honest historians.
In mathematics education I called the modern thought systems that we get as helping platforms the "silver platters". The most famous is the decimal number system of course. The point is not the decimal rather the use of zero that allows the bigger and bigger exponentially growing groups. For natural numbers this is merely practical as short expression of numbers. But when we continue the idea as division by bigger and bigger powers of the base like the tenth, the hundredth, and so on, then we get the infinite decimals. Here, the length is thus approached as an infinite sum at once. So, we instantly get the resolution of the whole Greek mathematics. A crucial point that is not emphasized in schools is that fractions when are actually divided with the digital methods, must always produce infinite decimals with repeating periods. Simply because the possible remainders are finite many. This fact then implies at once that not all decimals are equal to fractions that is rational. All non periodic infinite decimals are irrationals. This was a major challenge to the Greeks and they could only show artificial special irrational distances. So the infinite decimals as silver platter indeed should make us feel superior. But there is a big chaos here unexplored! Behind all these lie the line as a point set. This vision is actually a modern abstraction and is the third mentioned consequence of Cantor's Set Theory as it penetrated the 20^{th} century. How shocking this is, can be best seen by going back to Euclid. He used the crossings of circles and lines continually and so the explicit vision should have been that these are nothing more that point sets and when cross they simply share those points as elements. Why was this not expressed? Why didn't Euclid discover sets? Because the points on the line were regarded as a complex affair! Our modern set consciousness is simply able to avoid further questions. Because we are more obedient and less inquisitive! So the modern child will say, yeah when we cross two lines they share a common point as element. But then if we ask, how about the next points on the lines, he just says well they are not the same. We have to push the child to see that actually there are no next points. How the points are on the line is simply ignored. So the set abstraction is an accepted umbrella that seems to give a vision without really giving one. By today the word "set" penetrated all levels and fields as a kind of exactness yet revealing nothing special. Elementary school textbook speak about the set of the natural numbers rather than just about themselves as objects. Should we abandon this flapping about sets when we say nothing concrete? It will not stop anyway, so we might as well accept that a hidden cultural shift has taken place.

The good news is that Set Theory is screwed anyway. It will have to transform or come back again from the sky as did for Cantor.
Most surprising is that the failure of Set Theory has not affected its success in Logic.
And this is our subject now. The failure is completely related to Cantor's original goal to distinguish the infinites. The above mentioned third cultural use of sets will probably stay for a long time but I think eventually this will disappear too.
So we should forget the bad and go to the success in logic but I make a long detour to sketch Cantor's original insight and mention the failure:
Up to Cantor the infinites were regarded as a simple opposite of the finite.
So even though we see that the infinity of the natural numbers seems much less than the infinity of an interval, where the points are not even visualizable next to each other, this was not pursued as a meaningful problem. An interesting in-between situation is looking at not all points just the exact repeated dividers with a fix division base.
For base ten, the first nine dividers are: .1 , .2 , .3 , ... , .9.
The second ones are ninety in number : .01 , .02 , ... , .99 where we don't have the zero ending ones .10 , .20 , ... , .90 because these are the same as our first nine was.
Then we get nine hundred new dividers again, and so on.
These finite decimals are very similar to the possible naturals except that here we have zero start as possibility after the decimal but exclude the zero ending ones. So then actually we could exchange these and thus for any natural get an exact finite decimal as follows:
Put a decimal in front if it if it has no zero at the end. And if it has some zeros at the end then put those in front and use a decimal too. Then the naturals:
1 , 2 , 3 , . . . , 9 , 10 , 11 , . . . , 19 , 20 , 21 , . . . correspond to all finite decimals:
.1 , .2 , .3 , . . . , .9 , .01 , .11 , . . . , .19 , .02 , .21 , . . .
So the naturals with fix steps in-between them were directly squeezed onto the line being denser and denser. This shows that these visual situations might be misleading. The lucky exchange coincidence was also irrelevant. Indeed, if we ignored the zero ending naturals or the zero starting decimals, the point is still that the rest are a sequence. The main common feature that we see both at naturals and decimals is that we list in bigger and bigger groups as having more digits. And this trick can then be used for a more amazing fact. Namely, that not just the exact ten or other fix based dividers but all possible dividers together are a single sequence too. Of course, allowing all possible dividers actually means that we regard all fractions. Indeed, the base is a denominator and the dividers are all fractions with that denominator. As I mentioned above for the ten based decimal system, the fractions are the periodic infinite decimals. This is far from trivial and only visible through the division process because of the limited possible remainders. This is a must to be shown in elementary school because then we see at once that any non periodical decimal we form, will be a non fractional point on the unit interval. But now for a sequencing of all fractions actually we can ignore this whole periodicalness in a fix base and go directly for all fractions.
The trick as I said is going in increasing groups. And indeed, with one fix d divider we only have finite many divider points, namely the d – 1 many fractions with smaller nominators. So here is a single sequence containing all dividers that is fractions:

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

To really see how powerful this finite group method is, we can now list not just these normal fractions that are smaller than 1 but all infinite many fractions that fill the number line between any two natural numbers. Instead of the nicer forms as "mixed numbers" that writes the whole part in front of such fractions, we simply allow the numerators to be bigger that the denominator.
This at ones "kills" our method above, because now with a fix denominator we have infinite many possible numerators and so listing these we would never get to the next denominator value.
The trick is now to go not in increasing denominators rather in increasing totals of the denominator and numerator. Then within the groups we can go again in increasing numerators:

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

Looking at this sequence is enough to make someone fall in love with Set Theory and yet Cantor had vicious enemies from the start. They said it's all smoke and mirrors, not real math.
When I was young and heard about these early critics of Cantor, I was furious. That was about at the middle point of Set Theory's short hundred year history. And now I see things differently.
I still will show that the early critics were wrong because Cantor proved actual classical claims in new and much simpler way. But I have to admit that if hundred years ago someone had asked me where will all this new infinity distinction lead, I would have said that it will transform whole classical mathematics. And this did not happen! So in a strange way the critics "jinxed" Set Theory and became a bit right at the end. The only argument we can show to use Cantor's discovery in classical math is pretty much the original.
A week version of this argument is for the old problem of irrational points that I just repeated above again. The non periodical decimals are all non fractional, so obviously there are a lot of these on the unit interval. I mentioned also that the Greek mathematicians created artificial constructions to show this same existence because simply lacked the base number system. Now Cantor's new infinity distinction provides a third way to see that there have to be irrational points.
Indeed, we just showed that the fractions are a simple sequence. So if we could prove that the full unit interval is a bigger infinite that is not sequencable then this implies at once that some points can not be fractional. This in itself would be enough to regard this infinity problem as real but actually Cantor used it for a much harder new question that was solved only not much earlier.
This is a generalization of rationality. To be a fraction for x could also be regarded as having a solution for the $dx - n = 0$ equation with d and n being naturals. Indeed, x can be directly expressed as the fraction with n numerator and d denominator. So then allowing higher order equations but with all whole numbers as coefficients, we can ask if the roots would be special.
Obviously they don't remain fractions because already a second order can have irrational roots, namely all square roots of whole numbers will become such too. The higher order equations again trivially contain the $\sqrt[n]{m}$ roots as the solution for $x^n - m = 0$.
The possible root values for any equation, became called as "algebraic" numbers and now the question became if these would maybe exhaust all possible numbers. Everybody guessed that this shouldn't be the case, that is non algebraic numbers must exist but to demonstrate such became like a modern version of the old Greek chase for irrationals. Only a complicated and artificially created concrete infinite decimal could finally proved to be non algebraic. But amazingly Cantor's argument practically works instantly for algebraic numbers jus as easily as did for the rationals.
Indeed, all we have to show is that the algebraic numbers are a single sequence. And this is not surprising at all with our group idea. The only external knowledge we need is that a fix n-th order equation can only have maximum n many roots. Then if we can list all possible algebraic equations, the roots are listed at once too as groups. But to list the equations themselves is easy by going again in a total value. This total of course should involve not just the absolute values of the coefficients but the order too. Then the equations with a fix total can only be finite many.
The $x + 99 = 0$ or $x^{98} - 1 = 0$ or $2x^{55} - 43x = 0$ equations for example, all have 100 exact total. Incredible many such exist because distributing 100 into all possible sums and using negative versions for the coefficients too is a lot of variations, but still only finite many.
After seeing how this sequencability can be stretched almost without limit to newer and newer infinites, we might be curious what will be the miraculous argument that shows why the full unit interval is still not sequencable. This argument is actually a concrete construction too and not even directly concerns the full interval. Namely, for any list of infinite decimals we will create one that is not present in the list. This then indeed implies that all can not be listed.
So let a list of decimals be:

.3604 . . . , .2057 . . . , .0523 . . . , .5379 . . . , . . .

Lets regard the D decimal that has as first digit the first digit from our first number that is 3.

Then as second digit the second digit from the second number that is 0. Then the third from the third that is 2 , then the fourth from the fourth 9 and so on: D = •3 0 2 9 . . .
Nothing forbids that this D could be by chance exactly one of our numbers in the list.
But now we alter every digit in this D, for example, add 1 to all, and thus obtain:

D* = •4 1 3 0 . .
This definitely can not be in our list because it differs from the first in its first digit, differs from the second in its second digit, and so on.
The conventional form is to write the initial list not as we did rather under each other. Then these chosen digits are in the diagonal and so D itself is the diagonal decimal while D* can be called an anti diagonal and so the method itself became known as the anti diagonal argument.
The most confusing about this "miracle" that we just witnessed is that is had nothing to do with the actual interval. So, only the infinite decimals being the points, that is the mentioned heuristic representation of points as infinite sums gives the result for points.
Amazingly, we can alter the argument to the interval directly and not use decimals at all.
So we have a list of P_1, P_2 , . . points on an interval and we have to show a P on the interval that is definitely not in our list.
First we repeatedly trisect the interval, that is regard the base 3 dividers. But now we'll regard not these dividing points rather the intervals themselves. Narrowing down on an always trisected path means that we make left or right or middle choices. In an other sense this is actually the tertiary, that is base 3 representation of the points but we don't care about this. Our only assumption is that any infinite L , R , R , M , L , M , . . . choice sequence will create a sequence of intervals within each other and it will contain a common point as the bottom or limit.
Now we can easily give such interval sequence that avoids the P_1, P_2 , . . point sequence.
We simply choose as start that third of the full interval that avoids P_1. So we see now why we needed trisection. Halving would allow that P_1 is by chance exactly at the middle an thus be in both intervals, but with three intervals, one must avoid P_1. Then we again choose a third length within our previous that definitely avoids P_2. They might all avoid it but that's immaterial, the point is that one third definitely must. Continuing this way, all points are avoided.
The use of the trisected fix subintervals was merely a nice convenience! Not only these must contain a common point but any narrowing sequence of intervals. Thus a narrowing interval sequence that avoids the P_1, P_2 , . . sequence can also be made by choosing more free intervals step by step. So, as start, there is an I_1 interval that avoids P_1 then there is one I_2 inside I_1 that avoids P_2 then an I_3 that avoids P_3 and so on. Since we always chose inside thus all previous ones are avoided too plus the intervals have at least one common P point.
This avoids all points of the sequence, in other words is simply different.
Cantor not only cared about the new non sequencable infinite he discovered as the total point set of an interval but also about how this point set looks inside. The above used common point axiom was one of his results. So, he became not just the father of Set Theory but Topology too.
In spite of this, the infinity distinction is the basic step and it begged the obvious question whether this new bigger infinity hides more such infinites or only one. To put it an other way:
Is there further infinite between the sequencable and the full interval? Or even more concretely:
Is there a point set on an interval that is more than a sequence but not as much as the full interval?
The "more than a sequence" means simply that any sequence formed from it will leave out members. But this "formed" actually meant using subscripts and this can be said more precisely as assigning unique natural numbers. Indeed, P_n is actually a (P , n) in short version. But this is important to clear because the other side that is not being as much as the full interval can only be said precisely this way. We don't have the naturals there, so subscripting is not even a usable custom there and simply we claim that there is no way to form unique pairs between our set and the full interval. This unique pairing is what we earlier called an isomorphism but that includes keeping all visible relations too. Here we don't care about any relations like how the point are ordered and so on. This intentionally hollow isomorphism is called an equivalence. The weird

thing is that every time we create equivalences we use some relations and so we never really create pure equivalence. But here we use the concept negatively. The full set of points being more than an assumed smaller subset just requires that equivalence is impossible. So the "being more" concept of sets has this totally negative and fictive feature inherently. Probably that's why nobody captured it before Cantor. For finite sets, the being more has positive verification. It is true here too that seven apples are more that four pears and so there is no way to assign them to each other, but here the four pears assigned to four apples from the seven will leave three there and that shows the impossibility of perfect pairing too. At infinites, a leftover means nothing. We can always create leftovers. For example, all naturals can be assigned to the fifth next and then the first five will become a leftover. Or even more drastically, every natural can be assigned to its double and then we only get the evens and so the odds become a leftover. This weirdness of the infinites was recognized by people before Cantor, amazingly by Galileo too. But this just left these in the false belief that therefore infinites are not comparable. The crucial new and totally negative approach of no possible full equivalence as distinction was the breakthrough of Cantor.
So back to Cantor's question whether there is such in-between set that is more than a sequence but less than the full interval, the tragedy is that we still don't know! And that makes the other direction, the topological examinations of the interval as point set very suspicious. How could the structure of the continuous line hide this secret. Is it possible that we still don't know how the points actually lie there? Cantor's own guess was that there is no in-between infinity and it became known as the Continuum Hypothesis. Which is a confusing name because it really doesn't concern continuity directly.
The real blast happened in 1962 when Paul Cohen published his proof that the axioms of Set Theory can not decide the Continuum Hypothesis. So, is this supposed to be in limbo? Why?
Or we just need new axioms for sets. Or as I believe, the whole concept of sets must be altered.
We should give more details of what Cohen actually proved because it is very surprising.
He showed, that even assuming not just one in-between S point set on the interval but a whole system of in-between sets that looks like :

$$\text{sequence} < S_1 < S_2 < \,.\,.\, < S_p < \text{full interval}$$

will not lead to contradiction. So these S sets are just some ghosts that we can not refute.
It's interesting why we placed an S_p here. It is meant as a directly previous infinity to the full interval. If we assume such then the other S can be truly arbitrary. But strangely, if the full interval would have no previous infinity then we know some restrictions that apply. Namely, the simplest situation that a sequence of bigger and bigger sized $S_1 < S_2 < \,.\,.$ subsets would give as total the full interval, that is $S_1 \cup S_2 \cup \,.\,. = \text{full interval}$ is impossible.
This follows from some beautiful results of already Cantor and his contemporary, König.
The start is a surprise already. The points of a square are the same infinity as an interval.
The unit square is of course simply all possible pairs of infinite decimals by using a Descartes coordinate system. And then the two coordinates can be combed into a single decimal's even and odd digits. From this we can feel how any finite dimension is obtainable. But the infinite dimensional "cube" remains the same infinite too. This has as points the possible sequences of decimals. Here a "combing" is more tricky yet surprisingly we already used this combing to make the double infinity of the fractions to become a single one. Remember, that going with nominators didn't work because for every denominator we have already infinite many:

$$\begin{array}{l}
\frac{1}{1}\ ,\ \frac{2}{1}\ ,\ \frac{3}{1}\ ,\ \frac{4}{1}\ ,\ .\ .\ .\ .\ . \\
\ \ /\ \ \ /\ \ \ /\ \ \ / \\
\frac{1}{2}\ ,\ \frac{2}{2}\ ,\ \frac{3}{2}\ ,\ \frac{4}{2}\ ,\ .\ .\ .\ .\ . \\
\ \ /\ \ \ /\ \ \ / \\
\frac{1}{3}\ ,\ \frac{2}{3}\ ,\ \frac{3}{3}\ ,\ \frac{4}{3}\ ,\ .\ .\ .\ .\ . \\
\ \ /\ \ \ /\ \ \ /
\end{array}$$

The trick was going in increasing sums which actually meant that we walked through this table by the finite diagonals marked above.
Now if we place the infinite many decimals as coordinates under each other, then again this walking through will create a single decimal for all such infinite dimensional points.
This simple fact that a sequence of sequences is still just a sequence also shows that the situation of $S_1 \cup S_2 \cup \,.\,.\, =$ full interval with all S_n being a sequence is impossible.
But back to why it is also impossible with $S_1 < S_2 < \,.\,.\,$ it's enough to show that in this case the set of sequences from $S_1 \cup S_2 \cup \,.\,.\,$ will be a bigger infinity.
$S_1 < S_2 < \,.\,.\,$ implies that if we assign to every point of S_n some points of S_{n+1} then some must be left out. And this also implies that if we assign to every point of S_n some infinite point sequences then in any fix positions in those, some point of S_{n+1} must be left out.
So imagine any assignment of sequences for all points in $S_1 \cup S_2 \cup \,.\,.$
Then regarding the left out ones for each S_n in the n position itself, gives a $P_1 , P_2 , \,.\,.\,$ point sequence. Observe that P_1 was left out from S_2 then P_2 from S_3 , and so on.
This $P_1 , P_2 , \,.\,.\,$ could not be assigned to any of the points. Indeed, to all in S_1 we couldn't have P_1 as first then to all in S_2 we couldn't have P_2 as second and so on.
As we see we have a use of Cantor's "diagonal" trick with a twist of assumed increasing sets.
This was discovered by König.
Back to Cohen's result, we have to mention that our oversimplified claim above that these possible in-between sets are just ghosts was misleading. Remember how enthusiastic Bolyai was about his new world. The non crossing yet approaching lines are ghosts too but have very rich particular features and here too, different assumptions lead to different features. But never to a contradiction. The huge difference is that in Geometry we have the perfectly plausible parallelity axiom. Even its most overcomplicated forms like Euclid's original was, or even the full meaning behind as the unlimited proportionality claim, are all plausible. Here in Set Theory we do not have such plausible meaning that would avoid or prove infinities in-between the sequence and the interval. We are missing something simple about sets! But some refuse to admit this and rather want to pretend that we have all cards on the table already. Then bend the truth even more and actually lie about the already existing plausibilities. In these plausibilities the question of whether there is in-between infinite or not, is a simple black and white choice. But these smart arses want to hypnotize us that actually it's not black and white, we do not have a "well defined problem to start with". The systems of plausibilities are driving mathematics. New wider systems emerge from puzzles in the old. To deny the legitimacy of the puzzle is a self deception to avoid the pain that I am so stupid that can't see beyond my present plausibilities to see the wider ones. And yet all these plausibilities are a priori lying in all of us dormant. So a visitor from the future could enlighten us instantly just as we could the enlighten the old geniuses who died in the pain of not conquering their own plausibilities. The advance of matter, the evolution of the material universe is a side effect of this deeper evolution. And the belief in the existence of a deeper evolution is Idealism. Talking about Idealism, we must mention how this whole puzzle of the sequence and the interval became generalized in axiomatic Set Theory. It gives a new dimension to this concrete case and raises the possibility that maybe only the general abstract form would lead to the new plausibilities. The start with the sequence as the simplest infinite is replaced by any S set.
The interval or the infinite decimals corresponds to the B^S exponentiation where the B base is the possible "digits" we used in decimals and now we again assign these to all S elements.
So the elements of B^S are all possible functions that are defined on S and take up values in B.
The simplest case $\{0 , 1\}^S$ contains all possible functions that use as value 0 or 1. We could use the {yes , no} as base too and usually we just abbreviate these binary functions as 2^S.
Just as at decimals the base was not important here we have the same but much more exactly. Here again, already 2^S is a set bigger than S and increasing B from 2 to practically anything is useless. This "practically anything" has a precise meaning because a trivial increase would be if

we use as a B anything bigger than 2^S. Then obviously, B^S is bigger than 2^S too. Now the amazing thing is that up to 2^S, all bases give just 2^S so: $(2^S)^S \sim 2^S$ where ~ means equivalence. This corresponds to the old result that the infinite dimensional cube has the same many points as the interval. Which meant that the sequences with merely binary two choices are same many as the sequences with choices as points.

König's generalization of the Cantor diagonality is generalized perfectly too. So the 2^S jump from S can be generalized to make a jump with assuming infinite many jumps already. These now don't have to be in a single sequence rather we assume that there is a k König function that assigns to some s sets always a smaller k (s) set. Then the total of the k (s) sets is also smaller than the possible choice functions that we can create by picking one element from any s. In formal notation König's Theorem is:

$$k(s) < s \quad \rightarrow \quad \bigcup k(S) < \prod S$$

Where now S is the set of all the s sets, k (S) is the set of all k (s) values, $\bigcup k(S)$ is the combined set of all these and $\prod S$ is the set of all possible choice functions from S, that is the possible functions that pick from every s one element as values.
This revealed a crucial point that only came out in axiomatic Set Theory. The use of choices.
Old fashioned naïve set theory uses the collection principle. This forms well determined sets, namely for all imaginable P (x) properties we collect all such x-es in the { x : P (x) } set.
This had its own problem because with some P (x) properties we get into contradictory sets and so this method had to be pruned leading to the allowable set collections as axioms. But to use random pickings is a whole new ball game. The Axiom Of Choice which simply claims that $\prod S$ is not empty, became the main concern not just because of König's Theorem but even more because it turned out to be unavoidable even to use the basic Cantor comparison of sets by equivalence applicable to all sets. I will come back to this application.

But for example the Cantor jump itself is not dependant on this axiom because for 2^S the existence is clear without the axiom. In { 0 , 1 } or { yes , no } we can simply regard one of them and so we have a choice with only these. In B^S we can still say: "pick one element of B and use that as fix choice" even if B is not concretely given. Indeed, this "pick one element" is merely a logical step. But in $\prod S$ we pick from infinite many arbitrary s sets. This is just as plausible as making choices from a single B but it can not be "exemplified" and thus its existence doesn't follow from logic. So as we see the Axiom Of Choice should be called the Axiom Of Independent Choices and it is actually a new logical axiom.
There is absolutely no plausibility that would doubt that we can make such choices but some amazing consequences by smart constructions lead to "weird" results or "paradoxes". None of these are paradoxical due to the choices, solely due to the constructions. And yet an idiotic belief spread that the Axiom Of Choice is not quite kosher.
The mentioned real reason of the problem that at arbitrary choices we don't have an example that verifies the existence at least, never occurs at the most typical set theoretical arguments that is at collections because if there is some x such that P (x) then the collected { x : P (x) } has this as element and so can not be empty automatically. But if we make not collections rather just descriptions about a desired set then it might very well be that we describe something totally plausible and yet we can not prove its existence. The prototype of this is the perfect counterpart of the independent choices.
The choices can be made simultaneously and instantly in space. The opposite would be to make choices one by one in time. But the real problem is just time itself and so the choices could be avoided or rather be dictated by a given f function. Starting from an s set the f function can grow by itself and only terminate when a set is reached where it is not defined. The growing could mean that for every already obtained new t we again find f (t). Unfortunately this is a very primitive method and only achieves s , f (s) , f (f (s)) , . . . as a possible total. So, merely a single sequence. The heuristic idea is to regard f being defined not on the last element rather on

the set of all the already achieved ones. So as start we regard not s rather the single set $\{s\}$ and ask if f is defined on this rather than on s itself. This is here merely a nuance with no importance. So suppose f is defined on $\{s\}$ and we have an $f(\{s\})$ abbreviated as $f\{s\}$.
Then the achieved set is now $\{s, f\{s\}\}$ and so we check f for this and if it is defined then the next achieved set will be $\{s, f\{s\}, f\{s, f\{s\}\}\}$.
The real beauty comes after infinite many new values when the total becomes :
$\{s, f\{s\}, f\{s, f\{s\}\}, f\{s, f\{s\}, f\{s, f\{s\}\}\} \ldots\}$
The possibility of continuation is now open because f could be defined on this set.

The precise concept of the continuation is the "f widening" $= S^f = S \cup \{f(S)\} = S + f(S)$ if f is defined on S. These are the wider and wider stages:

$$\{s\}, \{s\}^f = \{s, f\{s\}\}, \{s, f\{s\}\}^f = \{s, f\{s\}, f\{s, f\{s\}\}\}, \ldots$$

Beside the obvious problem that f is not defined, a new possibility is now that a widening can stop for an other reason too. Namely if the "new" $f(S)$ value is not new rather it is element of S already. To make the f widening always be defined is easy:
$S^f = S + f(S)$ if f is defined on S and $S^f = S$ if f is not defined on S.

And then the definition of an f terminating S is simply $S^f = S$.
But now this includes not just f not being defined but $f(S) \in S$ too.
We feel that an f starting from a particular s will widen or grow by itself till it reaches a first f terminating S and we could call this stage the growth of s through f. The set of all wider and wider stages together should be called as the growth of f from s.
Can we describe the growth of s ?
Unfortunately, the only sure thing we know that it must contain s and it must be f terminating.
The trick towards a better description actually starts with realizing an error we made.
The f widening is not enough to achieve all stages! Indeed, after infinite many stages we had to combine these and this was a hidden extra step. This means that for open beginnings we add the combined set of these beginnings as new stage. First of all, the open restriction can be neglected because if the beginning is closed, that is has a widest stage then it is the combined set anyway.
So the only problem is to describe what a "beginning" should mean.
If we target the growth of s , that is the first f terminating stage then unfortunately the stages all melt into this and so the beginnings would be special sets of subsets. If however we regard rather the G set of all the stages up to this stage, that is the growth of f , then the beginnings here in G are special subsets.
Then the chased first f terminating stage will still come out as the total combined set of this G , that is as $\bigcup G$. So we achieve the growth of s too.
Still, we need a description what B subsets of G should be the beginnings.
The outside of a B, that is $G - B$ is the E end and all stages there should be wider than in B :
$(S \in B \text{ and } T \in E) \rightarrow S \subset T$
This allows B or E to be empty or the full G but we have to exclude empty beginnings when we claim the combining. But allowing the full G as beginning implies that $\bigcup G \in G$ and the final fourth rule says that this is an f terminating stage.

$\{s\} \in G$

$(S \in G \text{ and } S \neq \bigcup G) \rightarrow (S^f \neq S \text{ and } S^f \in G)$

B is a non empty beginning of G $\rightarrow \bigcup B \in G$

$(\bigcup G)^f = \bigcup G$

Unfortunately, as I foretold, we have a hollow description!
We can not guarantee any examples of such G stage sets because we simply aimed too high for the unique total G and the existence of that is non trivial. The solution is simple.
We have to relax the concept of widening stages to partial W ones that obviously don't obey the last rule only the first three. Unfortunately, this is still not enough! Our rules were not just too strong but also too weak! Indeed, they only tell what must be in but do not exclude any junk.
So we need a new fourth rule to exclude junk:

$\{ s \} \in W$

$(S \in W \text{ and } S \neq \bigcup W) \rightarrow (S^f \neq S \text{ and } S^f \in W)$

B is a non empty beginning of W $\rightarrow \bigcup B \in W$

$(S , T \in W \text{ and } S \neq T) \rightarrow (S \subset T \text{ or } T \subset S)$

The final trick will then be to combine all these partial W widenings to get G.
We have obvious existence for W because already $\{\{ s \}\}$ becomes an example.
And so we have existence for G too as it now contains the W-s as subsets.
So the subjective growth as a timely concept becomes a spatial collection of W-s.
The start of both the exact proofs and the proper vision is the following fundamental fact:
For any B beginning of a W that is not the full W, the $\bigcup B$ or $(\bigcup B)^f$ is a new stage not in B rather in its end and this has to be a subset of all others in the end. In other words, this has to be the narrowest in the end. So as a consequence we obtained that in any end the common part or intersection of all stages is actually a stage in the end :

E is a non empty end of W $\rightarrow \bigcap E \in E$

This seems similar to our third rule but here we have common part, that is all common elements as set and more importantly this stage is claimed not just being in W but in E itself.
To see really why our fourth rule was successful to avoid junk, we now generalize this end rule to arbitrary C collection of stages:

$(C \subseteq W \text{ and } C \text{ is non empty}) \rightarrow \bigcap C \in C$

Indeed, lets regard all those S stages that are subsets of all stages in C.
This is a B beginning because any stage that is not in B must contain some stage in C and so it can not be subset of any stage in B. So we have a minimal stage in the W– B end and it has to be a C member otherwise it were a subset of all C members and so it would have to be in B.
An other way of saying this rule is that for any kind of stages there is a narrowest such.
So then if we had junk among our stages then there would have to be a narrowest junk J.
The impossibility of this follows from the simple fact that for any S stage in a W the stages that are subsets of S are a B beginning and S is $\bigcup B$ or $(\bigcup B)^f$.
So the sub stages of our previous J were a B too, containing no junk and J itself would become $\bigcup B$ or $(\bigcup B)^f$ which are then non junks.
Of course, this was a bit of external logic and not quite exact.
The exact proof for the existence of the original G first shows that a G as a combining of the W-s obeying the four new rules will obey these rules too. Then it's easy to see that the $\bigcup G$ total is f terminating and so actually the original fourth rule is obeyed by G too.
Indeed, G has to be the widest possible W but if $\bigcup G$ is not f terminating that is
$(\bigcup G)^f \neq \bigcup G$ then $G + (\bigcup G)^f$ were a wider W.

The crucial inheritances of the new rules to G follow from two simple facts:

Any closed B beginning in a W widening is itself a W' widening.

For any two W , W' widenings, one is beginning of the other.

What I just explained is the true spine of Zermelo's Wellordering Theorem.
So the heuristic G growth of the arbitrary f function from an arbitrary s start is the essence.
The G growth is trivially a wellordering of its $\bigcup G$ total if by wellordering we mean ordering where every beginning has a next member. But the important part of the Wellordering Theorem is that such ordering exists for arbitrary A set. So the obvious task is only to show that every A set can become a $\bigcup G$. And of course to get the suitable G we only need suitable f and s.
This is again a phenomenal trick but has no technical part in the time avoidance.
Here, we simply regard all possible S subsets of A , now including A itself too and make a c choice function by the Axiom Of Choice. This c picks an element from all subsets of A including the full A. In fact, this choice from the full A will be regarded as s the start and c on the real subsets will give our f as : $f(S) = c(A - S)$.
The beauty of this is that it is always outside S and so it can only terminate by not being defined. Which means that S has to be A. Indeed, every other S value gives an A – S that is defined.

The possible wellorderings are also "similar" to each other's beginnings, that is isomorph with the two ordering corresponding to each other. This then instantly implies that any two sets if wellordered can be compared by equivalence because the similarity is at once an equivalence too.
So as I mentioned, this proves that Cantor's comparison by equivalence is usable for all sets.
Of course just because one becomes a beginning of the other doesn't mean that it is a smaller set. It only means that it can not be a bigger. So the vision of the wellorderings doesn't really show the jumps in set sizes. The sequencable wellorderings just grow and grow and miraculously become non sequencable. The forward growth or wellordering has its simple rule that we expressed clearly by rule 3. that after every beginning there is a next.
In any ordered set a beginning and end pair has only four possibilities.
Namely, the beginning having last element or not and the end having first element or not.
The beginning having always next element means also that the end always has first. As we saw this generalizes to all subsets having first element. A trivial consequence of this is that a backwards going sequence $\ldots s_3 , s_2 , s_1$ is impossible. And this is a paradox because we have arbitrary big infinites going forward. Yet, backwards we can only go finite many times.
But observe that already in the natural numbers we have infinite many and yet if we pick one we'll have only finite many smaller. In wellordered sets we have the same limitation of only finite many possible picks backwards. This picking idea also shows why this special paradoxical feature of the wellorderings is actually a sufficient condition already. So it implies that all subsets must have first element. Indeed, if a C had no first element then we can just pick any element of C and we can be sure that there are ones before. Then pick again and again earlier ones and so we gat a backwards sequence. This was again a "timely" argument and to make it spatial we need the Axiom Of Choice.
The most interesting question is whether the plausiblity of the growth extends to a plausibility of the wellorderings. At first they seem to be just boring repetitions. Every continuation must start all over from the beginning. But for example achieving a perfect repetition of an α type we get the double $\alpha + \alpha = 2\alpha$ as new type. Then the triple and so on, leading to $\alpha + \alpha + \ldots = \omega\alpha$.
This then again starts from scratch and so actually we not just always start all over but inflate the earlier steps with using any α as a single step. But this newer boringness is misleading again.
We can always ask questions that can not be answered by any boring visions rather by going into the jungle and see how rich the possibilities are. For example, we can say that these $\omega\alpha$ jumps are big because any beginning in these is simply "cutable", that is the left over or the end remains the same type. So such types should be called as end types in general. But as easy to see, if an α is end type then the next is $\omega\alpha$ and so these end types are merely the boring ω multiples.

These $\omega\,\alpha$ "jumps" are obviously not jumps in sizes and interestingly they can be used to prove that even the $\alpha\,\alpha = \alpha^2$ types remain same sized, that is $\alpha \sim \alpha^2$. This is exactly so because this boringness of the end types means that every type is actually $n_1\,\alpha_1 + n_2\,\alpha_2 + \,.\,.\, + n_m\,\alpha_m$ that is a sum of finite many repeats of smaller and smaller end types. This allows to comb together any pairs of types under α to get a new type still under α and such pairs obviously give the types up to α^2. So we indeed get that $\alpha \sim \alpha^2$. The combing is simply to regard the sum of the two finite combining from end types. This $\alpha \sim \alpha^2$ then implies $S \sim S^2$ for any sets.
So the boringness of the end types was very useful. But then we can even ask a question that goes beyond the boringness of $\omega\,\alpha$ and tries to grasp a deeper meaning of "end"-ness.
Namely, whether a β could be such, that if any S subset of β goes all the way in β, then S would have to be the same type as the full β. The not going "all the way" is easy to define, namely as being "confined" to a beginning that is not full. So this confined simply means being subset. Then of course the S "going all the way" becomes precise too as S not being subset in any beginning that is not the full β. And this can be called as S being a "cofinal".
And now comes the important result that shows how our new concept of cofinal succeeded:
If after an α the first bigger infinity type is β then this has to be such that if it has γ type cofinal then $\gamma = \beta$. Indeed, β is γ many smaller than β that is maximum α sized sets combined. Now if γ were smaller type then it were smaller size too because we assumed that β is the first of its size. And then γ could only be maximum α sized too. Thus, the total combined set could only be $\beta \sim \alpha^2 \sim \alpha$ a contradiction.
So we got quite a good picture about the previously called "miraculous" growth jumps in sizes:
In these sets, no beginning can be stretched out to go all the way.
The reverse of this vision is also true! Such wellordered sets are always a first sizes, that is have no beginning with same size. Indeed, if that was the case, that is we had an equivalence with a beginning then we could actually stretch a subset of that beginning all the way. All we have to do is regard only those s elements of the set that have assigned t pairs in the beginning so that no t' before t is assigned to an s' after s. The first t_1 element of the beginning is obviously assigned to such s_1 for example. The second t_2 is not necessarily because it can be assigned to an earlier than s_1. But there has to come a t again that goes forward in s too because otherwise the elements after s_1 would not be assigned. Repeating this we get a sequence of t-s that goes forward but this is still not necessarily a cofinal, so we would have to argue to restart new sequences again and again. Thus only a "timely" argument could "justify" the cofinal.
The precise goes indirectly as follows:
If those s elements that are assigned to t that have no previous t' assigned to s' after s were not a cofinal, that is were in a non full beginning of the set, then there were an s_1 after this beginning and it would have to be assigned to some t_1 that has earlier t_2 that is assigned to a later s_2 than s_1. Then this s_2 were not in the chosen s elements again and so t_2 would have to have an earlier t_3 assigned to an s_3 and so on. We thus get a backwards infinite t sequence.
So here the repeat argument was okay because it relied on a single sequence!
Interestingly, the reversal was actually stronger than the original result. Indeed, there β was assumed to be a next infinity after an α. If it is a new infinity without previous then we know that sometimes it can have smaller cofinals but an interesting new jungle is the question whether such can be without smaller cofinal. This obviously relates to the Continuum Hypothesis.
Finally, a very different and perfect condition can be given for the question whether for an A set, and R set of its subsets, there is a B between subset in size, that is with: $A > B >$ all $C \in R$.
Let $[R]$ denote all subsets of A that are same or smaller sized than any $C \in R$.
If such B exists then and only then for any f function that is defined on A and has a range inside $[R]$, we have two $a, b \in A$ so that $a \notin f(b)$ and $b \notin f(a)$.
The easier direction is the then. Indeed, lets have a B and f that $f(s) < B < A$.

Let f (B) denote the combined set of the f (s) sets for all s ∈ B. This is still just B sized and so can not exhaust all elements of A and thus we can choose an a outside of f (B).
The same way, f (a) is smaller than B and so there has to be a b in B that is not in f (a).
So at once we get that b ∉ f (a). But also, f (b) is subset of f (B) and so a was not just outside of f (B) but also f (b) , that is a ∉ f (b).
An application of this direction is the following: If on an A interval we assign to every P point some finite many points f (P) = { X , Y , ... , Z } then there have to be two P , Q points that are not members in each other's assigned sets. Indeed, [R] must be chosen as all finite point sets in A and B as any sequence of points in A.
For the only then direction we must show that if for an A set and R set of its subsets there is no B between subset, then there is an f so that there is no a , b pair that exclude each other.
[R] now must contain all smaller sized subsets of A and thus to demonstrate such f is easy!
First we regard the minimal wellordering of A. This then has only smaller sized beginnings and so regarding as f (s) the beginning up to s we have the range of f indeed in [R].
Clearly, for every s there are elements outside f (s) , namely all the elements after s. But none of these will have an f value that excludes s because s will be in the beginnings of those.
So as simplest example, if among an infinite sequence of people everybody likes only finite many others, then this doesn't imply that there have to be two that don't like each other. Indeed, if all natural numbers like the smaller ones, then though there are many not liked by any fix n , we can not choose one from these that wouldn't like n. Of course, as we saw above, on the interval such finite valued f function does imply symmetrically excluding points. Allowing f to order sequences to any point and assuming that the sequence is the biggest infinity under the continuum, we get the same visual f example without symmetrical exclusion for an interval as for a sequence with the "likings".
Indeed, by looking at a minimal wellordering of our interval, all beginnings are sequancable and so we can define f as these before any point. For every P point the liked ones are before and the not liked are after and so those will all like P. So there won't be a symmetrically unliking one.
Chris Freiling tried to claim that regarding the interval as a randomly choosable set, justifies the symmetrically excluding point pairs. Aside from the problems within his arguments, in my opinion the plausibilities in our counterexample should be involved too. Indeed, the minimal wellordering is not doable randomly and the sequences for an f are neither as beginnings.
So a much wider rethinking of Set Theory is needed than merely an Axiom Of Symmetry.

An other seemingly wild and opposite direction to shed some more details on Cohen's result is the following: As we saw, the unsequencability was actually a final indirect twist by assuming that the total set is also a list which is impossible because all lists have points missing. Now what if the list argument is correct but the total is simply a more complex list for which we can not apply the argument. For example, the members in this complex list can not be altered diagonally because there is no way to grab them. So then the full interval is actually a sequence but not expressible as a sequence. And indeed, Cohen only regarded weird sequential realities for the whole Set Theory to show the no contradiction derivability. And this is not accidental! As we'll see, these sequential models are part of the very success story of sets as structures for all mathematical realities.
Models are the same as structures if we have an axiom system in mind.
As we just learnt, the axioms for sets themselves were shown to be insufficient by Cohen yet now I claim that sets for structures are perfect. It is still a puzzle today. But apart from this puzzle, to be truthful, axiomatic Set Theory is a nightmare even for that perfect goal, that is to be structures.
A good example is going back to geometry. I lamented about how the concept of points as elements of the geometrical objects like lines and circles became a silver platter vision. And indeed, modern education should use these, regardless of the ancient Greek thoughts about how the points are unimaginable. But most shockingly, a precise Set Theory can not handle points either. Points themselves have no elements and so their distinctions should be given as start. Instead we lie! We have the naïve set vision as acceptable and yet when we use concrete structures we cheat and avoid points. So we use some fix representation like above the narrowing intervals and use these as replacements for points. A lot of these problems culminated in Hilbert's

new axiomatic Geometry. He was probably the greatest admirer of Cantor and to prove his belief in sets and also in his own exact logical rules, he decided to fix up Euclidian Geometry. The number of axioms became much more and they became much more intricate. This new Geometry will never enter high school education which is a pity because it could be deformalized and would give an incredible insight into the true missing parts in Euclid's system.

Beside the over simplification that all objects should be just sets and thus no points are allowed, a further major "uglification" is that the classical mathematical realities with objects having some relations are actually represented as parallel ghost sets. A family is not merely the set of the family members but being brother is again a set that contains now the brothers as pairs. Then even this is overcomplicated because strictly the pair (J , P) is not allowed either. Only $\{ J , P \}$ that contains John and Peter but without any special order, so $\{ J , P \} = \{ P , J \}$. Luckily, then a trick can be used, namely to regard the two of them as orderless set but also one of them as a repeated outsider, in other word the set $\{ \{ J , P \} , J \}$ or $\{ \{ J , P \} , P \}$. Then these two are not the same and so can replace the ordered pairs (J , P) or (P , J) respectively.

These technical "tip toe"-ings are all part of real Set Theory but have no importance in the actual meanings that we would have naturally as sets with inner conditions, that is as structures.

So actually, the concept of "Structure" is an existing plausibility that we still never claimed just shoved under the carpet of naïve Set Theory.

The concrete start of logic is what I revealed already as Beltrami's observation, that the weird models actually proved that unique parallelity can not be a consequence of the simple axioms because we can see such models where those axiom are all true yet there is no unique parallelity.

And as I said the hidden assumption is that models obey logic, so if the simple axioms would logically imply unique parallelity then it would have to be true in these models too.

Without exactly knowing what is logic or being model, we were able to claim such a grand claim that models that is reality must obey logic. And so the natural question is this. Shouldn't there be a reversal too? Shouldn't logic obey reality too? It seems quite simple too how. Logic is the universal truth and so it is true in all realities. In reverse, all those truths that are present in all realities should be logical. A refinement of this is if we realize that logic is used to go from assumptions to consequences exactly as we used the obeying for the parallelity. So the simple axioms were assumed and we said that logical consequence would mean consequence in reality too. So then in the reverse obeying again, we assume some $A_1, A_2, \ldots, A_m$ claims and we only have to look at realities where all these are true and if we see that a C is always true in these then actually this is a logical consequence: $A_1, A_2, \ldots, A_m \vdash C$. Where this $\vdash$ symbol actually means that our concrete future logic must be able to derive C from the left ones.

Now a major problem is that sometimes we can have infinite many assumptions.

With the original direction this is not important. If reality obeys logic then any derivation that can obviously only use finite many of our assumptions will be true and so the infinite many will include these assumptions anyway. But in the reversal, we don't know which finite many assumptions are needed to get the logical derivation. So we would have to pick any finite many and check our C for these as derivable. This is too complicated! Luckily, we can simplify the situation by realizing a few things already about our future logic, The first is what we saw as start of the non Euclidian geometry discovery. It was using indirectness. So here again if we want to derive C then we might just assume the negative of C abbreviated as $\neg C$, add it to the assumptions and derive a contradiction: $A_1, A_2, \ldots, \neg C \vdash$ contradiction.

That didn't help because the special C is still there distinct from the assumptions. But lets regard what we really wanted. It was to have a condition when our logic obeys reality. The reality is simply that there is no reality with $A_1, A_2, \ldots, \neg C$ all true and then this should imply that a contradiction is derivable from some of these claims. So here the $\neg C$ can be just any of the assumptions already. Then all we want is that any $A_1, A_2, \ldots$ assumptions that are impossible together in all realities should be able to derive a contradiction from a finite subset.

The only ugliness is now that we don't just have to check all possible assumption sets but even all realities. And to avoid this is simply to regard the negative form of our claim. Indeed, every implication is reversible and for example fever implying illness also means that no illness implies

no fever. So in our case, impossible model implying contradiction derivation means simply that no contradiction derivation implies no impossibility of model, that is a possibility of some model.
The no contradiction derivability is called consistency and thus the simplest claim that means a complete logic that obeys reality is simply that:
Any consistent set of $A_1, A_2, \ldots$ assumptions must have a model.
The consistency of course means all possible contradiction searches without success and most amazingly these unsuccessful searches can indeed produce a model.
Of course this is a big surprise in itself! Because these searches are a sequence and thus we at once show that all assumption sets have simple sequential models even if they talk about big non sequencable sets. And indeed, as I already mentioned Cohen also used such sequencable models of Set Theory. Plus our final strange remarks about a complex sequencable total makes more sense now. The sequenced model we create from the assumptions is such external sequence not visible inside the model and thus can be claimed as non sequencable inside.
This weird model phenomenon also true in the opposite direction as having big models for small intended realities. But observe that these weird models do not blemish the perfection of our logic only cast doubt on our axiom systems. And indeed the second part will prove these doubts.
This previewed sketch of why our logic is complete was very important for an other reason too.
It shows most concretely that in meta mathematics everything is perfectly explainable didactically. Not only we can avoid ad hoc ideas in the proofs but a much deeper trap of ad hoc definitions can be avoided too. Meta mathematics can be built without tricky definitions that simply work in the end. And the basic rules of logic as start is not exception either.
Yet, every single Mathematical Logic book on our planet starts with a mish mash of intuitive and arbitrary assumptions as start. Then in the end the Completeness of the Logic is established as a miracle. But completeness is evident from the real logic we used way before discovering Logic.
The arbitrary rule parts of our logic is a consequence of the Completeness and not in reverse.
I mentioned a lot of plus and lot of minus but as final judgment stood up for Logic. So repeat this basic claim again! The logic we use for thousands of years became crystallized in Meta Mathematics and the negative parts that we can not describe realities perfectly is not a fault in our logic. Where exactly the fault lies is not clear yet. But to blame logic or introduce artificial, more "reliable" logics is joke. To investigate such alternate logics is everybody's private business and maybe even useful. But to deny the universal validity of THE logic we crystallized is a lie. The crucial concrete point is of course to stand up for indirectness as the basic way of arguing. In Cantor's negative concept of being bigger as having no equivalence then this negativity reentered again. So those who want to avoid indirectness and accept only constructive proofs would have to throw out set theory too. But the point is that these artificial logic restrictions are all artificial. There is no narrower logic that we use naturally.
And yet we have to emphasize that this logic I am defending is just the logic of derivations. Arguments to prove and not to understand. So my earlier anti Formalist explosions are still valid.
The logic of understanding or as I call it the "Didactical Logic" of the future is not altering this logic we crystallized in Meta Mathematics. They only will interrelate in some sense.
But instead of the future we should tell some historical facts.
The proof for the Completeness of the new logic came very slowly and was only presented by Gödel in 1930 at a congress where Hilbert was present too. Probably he couldn't follow the arguments fully but his belief in this Logic was confirmed finally. But he also had high hopes for the axiom systems we use too. He thought that we can decide all problems from axioms.
As we already emphasized the completeness of a Logic doesn't mean this, it has nothing to do with how good our axiom systems are. His expectation can be called a completeness of the axiom systems. The truly amazing fact is that in 1930 when Gödel presented the Completeness of Logic he already knew that this axiomatic completeness will not be true and talked about this in the intermissions. The young John von Neumann was one of the few people who could follow the thoughts of the young Gödel. A year later indeed the proof was ready, the well known axiom systems are not complete. So as we see, a single person's, Gödel's two theorems, the Completeness Theorem and the Incompleteness Theorem are the two major results of Meta Mathematics. The two completeness words of course are different. The first is about completeness

of Logic, and as we explained it boils down to the fact that all consistent assumption sets must have models. The other, the completeness of an axiom system means that all A , ¬ A statement pairs are decidable, one of them is a theorem.

Now back to the real start and explaining logic, we'll see that not only the derivational rules themselves can come out logically without ad hoc presumptions but the language of logic too.
I already mentioned negation ¬ A and implication A → B.
Historically, this implication was a falsely over emphasized operation. The reason partially is that a so called Formal Logic existed for two thousand years and this "half baked" logic had its basic rule using implication. Namely, it said that if A → B is derivable and we can derive A on its own too then we can infer B. In real life this is so obvious that we don't even need a rule for this.
More importantly, the real question behind this rule remains what kind of A → B implications should be derivable and how. Old formal logic didn't reply to this perfectly. The new logic of course went into the derivable implications but exactly due to clinging into implications, the solutions were complicated. It took some clarifications and simplifications to give up implication altogether and rather start with the two more natural operations "and" and "or".
Special symbols again can be used like A ∧ B and A ∨ B.
While in everyday language the "and" is perfectly clear, the "or" has three meanings.
The permissive "or" which we use as A ∨ B means that at least one of A or B is true.
"Permissive" meaning that both can be true too.
But sometimes we use short "or" for "either or" meaning which is abbreviated as A ∇ B.
Finally, rarely we use "or" for exclusion, that is meaning that at most one of A or B is true but allowing that neither is true and this is abbreviated as A | B.
Another frequently used symbol is A ↔ B meaning implication in both directions that is:
A → B ∧ B → A.
But these abbreviated everyday meanings for logical operations are still just a few of the theoretically possible fourteen possibilities. Indeed, if from A , B we create a combination then this combination is determined by how the truth of these two coincides with the truth of the combination itself. For example ∧ means that only true A and true B gives true ∧. So the truth list is { (t , t) } . For ∨ the truth list is : { (t , t) , (t , f) , (f , t) }.
Now, from the four (t , t) , (t , f) , (f , t) , (f , f) input possibilities we can make fourteen possible lists, namely four singular ones, six doubles and four triples.
The no input would be always falsity and the all inputs would be always truth. These two can be abbreviated as the F and T fix truth values.
Important to mention that the truth list of → is { (t , t) , (f , t) , (f , f) } so the only falsity of → stands for (t , f). So, false A assumption automatically means a true A → B implication.
So the claims that if the pope is a woman then there is God or there isn't God are both true.
Best is to regard A → B as merely an abbreviation for ¬A ∨ B or ¬(A ∧ ¬B).
All these were unimportant because the everyday meanings support everything easily.
The real heart of Logic is based on two operations that lurk behind everyday language so well disguised that people never reach consciousness of them without external push. And yet these two hidden meanings are not only rule the mathematical use of language but very important in everyday thinking too. So it is not an exaggeration that the revealing of these two meanings in our everyday languages should be a most important point in elementary school for its own sake as clarifying our thinking even in the most common matters. A second bless would be of course a preparation to mathematics, beyond numeracy! To open the joy of mathematics for all.
This analysis of everyday sentences that I call "grammatics" should be explored very early. It is the fourth absolute minimum after the three "R"-s, reading writing and arithmetic. This of course will never happen in the present mind set of education.
I will show with concrete examples how simple this Grammatics is and the revolutional consequences can be seen by anyone who is still fiery about truth. But this simpleness is actually misleading. The mentioned two crucial logical operations as magic wands only work with a deeper underlying magic, the real secret of logic. And indeed, the two operations themselves were discovered or collected by Aristotle already, but used in a totally false manner because the real

magic was unknown. That part only came out by the 19^{th} century through mathematics, even though the magic itself is not essentially mathematical.
The two formal magic wands are the everyday meanings of “every” and “some”.
Seemingly the “some” is much less than the “every” and yet this “some” is more interesting because it hides the “there is some” and so actually existence is claimed. All this sounds like some existentialist hocus pocus but soon it will make perfect sense. The best start is to realize that these two, the “every” and the “some” are related to each other by negation. “Not every person is good.” indeed means that there are people who are bad that is not good. So the “not” from the start went to the end. Also in reverse, not being true that “Some people are bad.” means that every person is not bad that is good. These are artificial “wisdoms” because they are obtainable by common sense anyway. In spite of this, these everyday meanings can already be useful to be spelled out before we apply them in math. And here I will reminisce a bit.
As I said I am very skeptical about the inevitable becoming the reality soon, that is Grammatics entering elementary school second grade where it should start. But this simple wisdom is kept hidden not only in high school but even in universities and even in math faculties, which is just simply a joke. It was a joke already in 1968 when I started uni in Budapest. I won many math competitions and got in uni without an entry exam. But I went to a new math high school not to one of the famous ones. My school was a miracle coincidence of an enthusiastic math teacher who turned out to be the best human being I ever met and a smart bureaucrat principal who was able to create a vibrant micro social environment for talents. At uni I met the competition face to face. At least ten of the students were from the famous high schools. As I soon realized they were not happy because the previous year was an exceptional content of talents from the special schools and so they got the most famous professors like Csaszar for analysis. The custom was to alternate yearly the lecturers and so our year got a “second grade” analysis lecturer by the name of Laszlo Czach who recently returned from Russia. This short bald guy wrote everything on the blackboard with incredible speed and so the girls and the Vietnamese students were so happy that the usual hunting for textbooks can be avoided. Everything was written down in full detail that was required to know. A trivial requirement from the educators that still is dodged in most unis.
At the first lecture Czach wrote down the following sentence:
“Every woman has a moment in her life when she’d like to do that’s not alright.”
It sounds much more rhythmical in Hungarian and as Czach explained, it was from an old song.
He said that this will be crucial to understand analysis which made everybody giggle.
Then he asked who could tell what the negation of the sentence is. Many didn’t know but soon we all agreed that it is: “There are some women who have no moments in their lives when they would like to do that is not alright.”
An S set of points approaches a P point if in every surrounding of P there is some Q point from S. This includes arbitrary small surroundings, so indeed S gets arbitrary close to P.
The opposite, that is S not approaching P thus means that there is a surrounding of P where there is point of S. So Czach fixed up the insanity of the education system as far he could and injected a bit of Grammatics into us as preparation. Of course I was in darkness at that time yet and only felt an instinctive appreciation for what happened.
There is a deeper layer to the story. My brother Peter was four years older than me and when I started math high school he started electrical engineering uni. His math lecturer was Csaszar’s wife and she wanted to show her husband how smart she was and wrote a totally abstract math text book. Peter couldn’t cope with the derivations and had to leave uni. I remember when I inherited his text books and went through the derivations. Peter even tried to find a tutor who happened to be Czach before he became a lecturer. My brother still became an engineer with many inventions. Csaszar’s wife is dead and knows nothing about me or my brother. Nor that Richard Feynman wrote a manifesto against assholes like her. When I read it, my memories flooded back and I realized first time what my poor brother had to suffer for no reason at all.
So now back to our real subject, I have to reveal why the “every” and “some” logical operators in themselves are not the real reason that Grammatics is so vital and what is the real hidden stuff in everyday languages that Aristotle missed too. The relations between objects that we use are very well expressed in everyday languages. For example I talked bout my brother Peter and so we can

say that there is a B relation that me John and my brother Peter are cases of this. So if $B(x, y)$ is the brother relation then $B(\text{Peter}, \text{John})$ is true. Of course, there are many Peters and Johns, so these to be used as names only work if we restrict these to us as concrete persons. Using the surnames is better but the phonebooks show that it isn't sufficient. The point is the intention in real languages anyway. When we talk about Napoleon we know which Napoleon we talk about and he was a concrete person. This writing concrete objects into the variables is the most obvious way of creating statements from "states". And state means in general not just relations but single variable "properties" like being good as $G(x)$ furthermore any combinations of such basic states by the already mentioned logical operations and negations. For example:
$B(x, y) \wedge G(x) \wedge \neg G(y)$ is the state that x and y are brothers, x is good but y is bad.
A concretization of this state is the statement: $B(P, J) \wedge G(P) \wedge \neg G(J)$
The really important other statement formation beside concretization is quantification!
Namely, by the two quantors every $= \forall$ and some $= \exists$.
But here we can not write these symbols into x or y because we can claim these for more.
For example, to claim that there exist brothers with one good the other bad would be stupid by:
$B(\exists, \exists) \wedge G(\exists) \wedge \neg G(\exists)$. Indeed this would mean that they are identical persons too.
We could use two different subscripted existence symbols as: $B(\exists_1, \exists_2) \wedge G(\exists_1) \wedge \neg G(\exists_2)$.
A more complicated yet later more useful method is to leave x , y in the state and tell in the beginning that they are meant as existence:
$\exists x\, \exists y\, [B(x, y) \wedge G(x) \wedge \neg G(y)]$. This is not only more complicated but seems stupid too because the order of the two quantification at the start is irrelevant. So it is the same as:
$\exists y\, \exists x\, [B(x, y) \wedge G(x) \wedge \neg G(y)]$. But this stupidity is everywhere and for example the three $\wedge$ members order could be altered too. The complicatedness is most visible by that we needed the square bracket because without it we wouldn't know if we quantized all the way or just the B relation. The real explanation for all this is that the initial order is actually very important and now was only irrelevant because both quantors were the same.
So for example the following two statements are totally different:
$\forall x\, \exists y\, [B(x, y) \wedge G(x) \wedge \neg G(y)]$ or $\exists x\, \forall y\, [B(x, y) \wedge G(x) \wedge \neg G(y)]$
The first says that for every person we can find an other so that they are brothers, the first is good and the other bad. An obviously false claim. The second statements says that there is a person that all others are brothers and the first is good while the second is bad. This is false again.
To say that there is some good person with bad brother is easy as follows:
$\exists x\, \exists y\, [B(x, y) \wedge G(x) \wedge \neg G(y)]$
To say that there is some good person with all brothers being bad is this:
$\exists x\, \forall y\, [G(x) \wedge (B(x, y) \rightarrow \neg G(y))]$
But this actually only said that all who are brothers are bad. So it didn't claim that there is such.
So to say that there is some good person having some brother and with all brothers being bad is actually to claim both statements above.
Here we see in action the existence feature and the weakness of implication as well.
In spite of all this, these meaning alterations by the quantification orders seem to be very natural to our subconscious and in my story Czach's fast introduction to Grammatics also relied on this.
So its quite a mystery why only in the 19th century became all this clear. There is only one explanation: The use of the quantification order is an innate or a priori ability exactly as Kant claimed many such, but the inducement from outside is necessary through the use of variables. This required the mathematical abstraction of variables in general and that was historically not ready yet even at Gauss. In my opinion Grammatics jumpstarts everything and prepares one for the mathematical variable use as well. So math should start only after Grammatics!
I show now three grammatics exercises I did for my daughter Timea when she was in elementary school:

1.) Timi's bicycle has been stolen. = Somebody stole Timi's bicycle.

$S(x, y) = x$ steals y

$P(x, y) = x$ possesses y

$B(y) = y$ is a bicycle

T = Timi

$\exists x \ \exists y \ [S(x, y) \wedge P(T, y) \wedge B(y)]$

2.) Children like candy.

$Ch(x) = x$ is a child

$L(x, y) = x$ likes y

$C(y) = y$ is a candy

$\forall x \ \forall y \ [(Ch(x) \wedge C(y)) \rightarrow L(x, y)]$

3.) If someone looked around on Mars, he'd see a strange view.

$L(x) = x$ looks around

$St(y) = y$ is strange

$S(x, y) = x$ sees the y view

$P(x, y) = x$ is on planet y

M = Mars

$\forall x \ \exists y \ \{ [P(x, M) \wedge L(x)] \rightarrow [S(x, y) \wedge St(y)] \}$

As a fourth example I want to detail Czach's sentence further than we did above:

Every woman has a moment in her life when she'd like to do that's not alright.

$W(x) = x$ is woman

$A(y) = y$ is alright to do

$L(x, y, z) = x$ would like to do y at z time

$\forall x \ \{ W(x) \rightarrow \exists z \ \exists y \ [L(x, y, z) \wedge \neg A(y)] \}$

Lets see the negative:

$$\neg\ \forall x\ \{\ W(x)\ \rightarrow\ \dots\ \}\ =\ \exists x\ \neg\ \{\ W(x)\ \rightarrow\ \dots\ \}\ =$$

$$\exists x\ \neg\ \neg\ \{\ W(x)\ \wedge\ \neg\ \exists z\ \exists y\ [\ L(x,y,z)\ \wedge\ \neg\ A(y)\ ]\ \}\ =$$

$$\exists x\ \{\ W(x)\ \wedge\ \forall z\ \forall y\ \neg\ [\ L(x,y,z)\ \wedge\ \neg\ A(y)\ ]\ \}\ =$$

$$\exists x\ \{\ W(x)\ \wedge\ \forall z\ \forall y\ [\ L(x,y,z)\ \rightarrow\ A(y)\ ]\ \}$$

There is somebody who is a woman and anything anytime she would like to do is alright.

Grammatics and my claim above that this sequential quantification we use at present is an a priori ability does not mean that other a priori visions can not attack it rightfully. Plausibilities have holes in them that can be seen by other plausibilities. That’s how paradoxes are resolved.
Here a major false vision is that the longer and longer alternating quantors are harder and harder to fill with meaning and so we think that they are becoming unimaginably complex. Yet there is something very simple common in all quantifications that is totally missed by the sequential.
Namely, the universal quantors can be claimed all together!!!!
So, in $\exists u \forall v \exists w \forall x \exists y \forall z$ the three universalities of v , x , z are a simple common universality and the u , w , y existences are only that distinguish them. To see this crystal clearly, we should introduce a new quantification as start and see how it can grasp the old sequential too.
To see everything even better, we should adopt a “tuple vision” first. Tuples are simply the finite sequences and so the (u , v , w , x , y , z) variable tuple as reality would be all those concrete tuples that can be formed in a reality by using concrete objects for the variables.
The best is to envision these under each other:

(u , v , w , x , y , z)

(1 , 8 , 657 , 9 , 79908 , 1)

(1 , 56 , 8 , 9 , 7 , 1)

.

.

The order of these concrete tuples under each other is irrelevant and they can be arbitrary big infinities not just a simple sequence anyway. This form is still good because we can see the columns under the distinct variables and this helps to see what a possible quantification actually means. The common universality of u , v , w then means not just that all concrete object of our reality must appear under these but that actually all possible combinations must occur.
The really big question is what we claim about the other variables.
This can be two totally different meanings of existence.
The independent would mean that there is an object under it and already with that concrete object we can find all combinations under the universalities. This would be claimed as a prior existence before all the universalities in our sequential method, like u was above.
These old frontal existences are the only true existences and they are simply a generalizations of using concrete names. Indeed with concrete name it’s trivial that with that fix name we have all universal combinations. Now such true existence simply says that we could have one such fix name but there might even be more such fix ones too. So it is simply a name avoidance.
All the other existences are completely different in meaning and they claim only that for certain universally appearing sub combinations we must have a common object in its column.

So now these should be not marked with $\exists$ rather with which universal ones should be sub universal. A possible situation then for the v , x , z being universal and u being unconditional could be the following:

$$\begin{matrix} & \exists & & \forall & & (v) & & \forall & & (v,x) & & \forall & \\ (& u & , & v & , & w & , & x & , & y & , & z &) \end{matrix}$$

The meaning of this is then that w is dependant on only v that is for every v there is a fix w(v) that will appear in the tuples. This of course means that actually this dependent existence is a universality claim because this w(v) will be universal for all x and z combinations. Similarly, y(v,x) will be universal for all z choices. This example actually gives the original sequential quantification $\exists u \forall v \exists w \forall x \exists y \forall z$ and so we merely showed that using w(v) and y(v,x) functions in places of w and y and a U name in place of u, we could avoid all quantifications. We could just use the state with all the other variables meant as universal. But there is an other twist to this method because we could have claimed different dependences as we did. For example, w(v,z) and then we would get not our sequential meaning at all. So this method seemingly opens new possible statements as well. But this is not really that important because these new statements can be replaced by more old sequential ones. The really important new insight was seeing that the common parallel universalities are the fundamental in all statements.
To prove how useful this tuple plus functional quantification vision is, we can show the earliest big surprise of logic. It is true without any actually logical rules yet and merely concerns structures. Namely, in any structure where a tuple set of some variables obeys a quantification, we can find a mere sequence of tuples that already obeys it too. We can very easily show how to choose this sequence and bring it up to the top of our tuple list. To be general we can assume that there were some $N_1 , N_2 , . . . , N_m$ names in our state too and these have already established objects corresponding them in our structure. Bring these up into their places. The rest are all variables that we have to fill too. These variables all have their quantifications above them.
First check out the independent true existences and find $E_1 , E_2 , . . . , E_n$ objects for them under the tuple list. Bring these up too. These N-s and E-s are the bare minimum that we must have in our topping, so we have to make all possible combinations from these under the universal variables but leave the dependent existences yet empty. Then, the reality, that is our list of tuples underneath will give examples for what those dependant values can be and we bring them up too.
The only problem is that this means new $O_1 , O_2 , . . . , O_p$ objects to appear and these were not used as combinations under the universalities. So we fix this by creating all missing ones and again bringing up suitable dependants. Repeating this, we get more and more imperfect finite tuples and we always try to fix them up. Amazingly, the full infinite many brought up tuples will become a perfect reality in itself. Indeed, the claim itself involves only finite many variables and so every chosen combination of the universal objects will have the required dependant existences there among the infinite many.
The argument involved a single statement and so we might think that using more or actually infinite many statements avoids that such primitive sequence from a structure could replace the whole structure. But this is not so. We can imagine now our sequence of statements one after the other and underneath each the corresponding tuple lists. We can then start our previous method but we have to use in our first state not just its own names and existence names rather the next one as well. Then we bring up the dependant existences and move forward in our statements too.
As new objects appear we use those too in all earlier states. Now much more combinations must be tried out as earlier for a single state but still only finite many up to any depth and going forward up to a state. So, the finally used objects from the reality is again merely a sequence and it will create a sequence of tuple sequences underneath each statement. And each again will be a perfect reality for each statement.
This so called “dovetailing” method was actually an other way of seeing that a sequence of sequences is merely a sequence but with some manipulations involved as well.

The result is nowadays called as Downward Lövenheim Skolem Theorem, meaning that a much smaller yet equivalent sub model was found.
The amazing thing is how early this was discovered and it didn't lead to the immediate discovery of the Completenes Theorem. With our present vision it is almost a single step. Indeed, the logical question is whether instead of using the already given reality, we could manufacture one.
Trying out the possible universal combinations could again be the start but now we don't know what dependences to bring up.
Best to jump into an example:

∀	∃	∃	∀	
1	∃	∃	∀	} ∀ -namings using 1
1	2	∃	∀	} ∃ -namings = making new names from 2 , 3 , . . .
1	2	3	∀	
2	∃	∃	∀	} ∀ -namings using 2 , 3 in old ones and 1 , 2 , 3 in new ones.
3	∃	∃	∀	
1	2	3	1	
1	2	3	2	
1	2	3	3	
2	4	∃	∀	} ∃ -namings = making new names from 4 , 5 , . . .
3	5	∃	∀	
2	4	6	∀	
3	5	7	∀	

4	∃	∃	∀
5	∃	∃	∀
6	∃	∃	∀
7	∃	∃	∀
1	2	3	4
1	2	3	5
1	2	3	6
1	2	3	7
2	4	6	1
2	4	6	2
2	4	6	3
2	4	6	4
2	4	6	5
2	4	6	6
2	4	6	7
3	5	7	1
3	5	7	2
3	5	7	3
3	5	7	4
3	5	7	5
3	5	7	6
3	5	7	7

∀ -namings using 4 , 5 , 6 , 7 in old and 1, 2 , . . . , 7 in new.

4	8	∃	∀
5	9	∃	∀
6	10	∃	∀
7	11	∃	∀
.			
.			

∃ -namings = making new names from 8 , 9 , . . .

Now I explain what's above. First of all, the process goes on up to infinity. I merely stopped after six such so called naming groups. As we also see, these happen in alternative groups as universal and existential and our case started with universal because this was the first quantor in the start. This reveals that the start or first line is a sequential quantification. Returning to these is the basic trick that allows to use the Lövenheim Skolem Theorem idea for a yet non existing structure.
The alternative namings for the quantors is actually a new layer of dovetailing inserted.
In our example we used the natural numbers as names. Some of them can be already names in the state that our quantification precedes. Above there was only one such assumed name 1.
That's why the first $\forall$-naming group had only one member.
If we had no names but lets say we had started with an $\exists$ quantor then we still would have introduced 1 as name for this and then we would have continued the same way as above.
The fundamental difference between the $\forall$-namings and the $\exists$-namings is that we return to all earlier $\forall$ quantors in the $\forall$-namings but we only use the $\exists$ quantors of the last group in the $\exists$-namings. That's why I even told in the $\forall$-namings what were the new names that had to be used in older, that is not last formed $\forall$ quantors while all the names up to the point had to be used in the new $\forall$ quantors. If more same quantor repeats then we can do these in one group as above for the two $\exists$. If two $\forall$ had been the case then it means all possible pairs.
This system achieves that up to a point we have always new names for all existences and more importantly we also have all possible combinations of universalities of the old names.
The situation is the same as was at the Lövenheim Skolem Theorem. The always occurring new existences screw up the perfect universality and we have to fix it up again and again.
And most amazingly, now it again works for the full infinite list. Every partially concretized statement will actually be truthfully further concretized. Meaning that all the existences will have names as case for them and all universalities will be exemplified by all possible names.
This then has an amazing consequence for the two extreme meaning of this partial concretization. One extreme is the original statement that has all original quantors and no artificially introduced name yet. The other extreme is all those lines that have only numbers. These are actually fully concretized statements. These two ends also perfectly correspond, meaning also that if we regard these totally concrete statements as definitions of a reality or structure skeleton, then the original statement will be true in that. Of course again we can extend this method to an infinity of statements and obtain a reality skeleton where all quantifications are true.
We had to say skeleton here because these concrete statements are not telling yet the basic states that is relations, properties or functions we want in our reality, only the cases of the states that we used in our fully frontally quantized statements. In fact we might even protest the assumption of such oversimplified claims only, that have all quantors in the front.
Luckily this is the easy part. Every statement can be easily transformed into such "prenex" form.
All quantors can be brought to the front. Of course then the state after the quantors can still be a very complicated expression even with containing only $\wedge$, $\vee$, $\neg$.
The second phase toward the Completeness Theorem is now relying on a beautiful further simplification of our states too. These are the state matrixes that won't even use $\wedge$ and $\vee$.
This can be done because first all the negations can be pushed forward to the basic states and then the remaining $\wedge$ and $\vee$ can be also reduced to one big single $\wedge$ of $\vee$-s or $\vee$ of $\wedge$-s, with the inner ones containing only basic or negated basic states.
This second choice, that is an $\vee$ of $\wedge$-s would look like:

$$(C_1 \wedge C_2 \wedge . . .) \vee (D_1 \wedge D_2 \wedge . . .) \quad . \; . \; .$$

Where all these C , D , -s are B basic states or $\neg B$ negated basic states.
And this can then be abbreviated as the state matrix:

$$\begin{bmatrix} C_1 & C_2 & \cdots \\ D_1 & D_2 & \cdots \\ . & . & \\ . & . & \end{bmatrix}$$

As an example of how a natural, that is loose statement can be transformed to prenex martrix form, lets regard the classical recognition of Euclid that there are infinite many primes.
As language, we need the naturals as names plus equality, smaller bigger and multiplication as basic relations: $x = y$, $x < y$, $x \bullet y = z$.
The composites are numbers that can be written as products without using the trivial 1 number.
So y is composite if $\exists u \exists v \ (u \neq 1$ and $v \neq 1$ and $u \bullet v = y)$
This means that 1 itself is not composite but it's merely because it's too simple. The other non 1 numbers that are not composites are the primes: 2 , 3 , 5 , 7 , 11 , 13 , 17 , 19 , . . .
The infinity of these means that they must become arbitrary big, and in reverse if they are arbitrary big then there must be infinite many, so our claim can be said by:
For every number there is a bigger non composite. Or in formal language:

$\forall x \exists y \ [x < y \ \wedge$ y is not composite $] =$
$\forall x \exists y \ [x < y \ \wedge \ \neg \exists u \exists v \ (u \neq 1 \ \wedge \ v \neq 1 \ \wedge \ u \bullet v = y)\ ]$

So we encounter the pushing in of the negatives, through the quantors which we mentioned already as Czach's example with naughty women for the limit.
In this case, the non existence of u and v means that for all u and v the composition fails:

$$\forall x \exists y \ [x < y \ \wedge \ \forall u \forall v \neg (u \neq 1 \ \wedge \ v \neq 1 \ \wedge \ u \bullet v = y)\ ]$$

Of course, then it's easy to push in the negative further by changing the $\wedge$-s to $\vee$-s :

$$\forall x \exists y \ [x < y \ \wedge \ \forall u \forall v \ (\neg u \neq 1 \ \vee \ \neg v \neq 1 \ \vee \ \neg u \bullet v = y)\ ]$$
The $\neq$ is already a negated $=$ and double negation cancels, so we have simply:

$$\forall x \exists y \ [x < y \text{ and } \forall u \forall v \ (u = 1 \ \vee \ v = 1 \ \vee \ u \bullet v \neq y)\ ]$$

The quantifications of u and v can be brought to the front:

$$\forall x \exists y \forall u \forall v \ [x < y \text{ and } (u = 1 \ \vee \ v = 1 \ \vee \ u \bullet v \neq y)\ ]$$

We only have $\wedge$-s and $\vee$-s, but in the wrong order. The $\wedge$ should be inside.
But that's easy to change because a first claim $\wedge$ some other's $\vee$ is the same as repeating the first claim for each $\vee$ claims again and again. The first claim is here $x < y$, so we have:

$$\forall x \exists y \forall u \forall v \ [(x<y \ \wedge \ u = 1) \vee (x<y \ \wedge \ v = 1) \vee (x<y \ \wedge \ u \bullet v \neq y)\ ]$$

This finally in a matrix notation is:

$$\forall x \exists y \forall u \forall v \begin{bmatrix} (x < y) & (u = 1) \\ (x < y) & (v = 1) \\ (x < y) & (u \bullet v \neq y) \end{bmatrix}$$

We show an example for this "logic machine" without containing any $\wedge$ and thus making the process without branchings much simpler.
Every time we encounter a new quantor we at once write a new variable next to it that will be used later as replacement for the old variable next to it. This makes the process really easy to carry out. Next we use the naturals as new variables and so we just keep increasing these quantor numbers. That goes for both quantors but the big difference is that at renaming the $\forall$ quantors they disappear and so the number increase will apply to a new encountered $\forall$. The $\exists$ application also makes this quantor disappear but we copy it to the end with increased number. On the other hand when we encounter a new $\exists$ quantor then it starts from start that is 1. So a lot of these existences will have different numbers. The method guarantees that all existences go through all possible values.
In theory this would never halt just go on forever. The beginning of course will have more and more basic relations without quantors. This is the decider of when to stop. Namely when we can see two exact opposite cases of a B and $\neg$B pair. This means a trivial truth and makes the whole line true. In the example next I pretended that I missed such pair and went two steps further that necessary. Indeed, $\neg B(1,2) \vee B(1,2)$ already appears in the sixth line:

$$\forall x \exists y \forall z [B(x,y) \vee \neg B(x,z)]$$

$$\forall 1\, x \exists y \forall z [B(x,y) \vee \neg B(x,z)]$$

$$\exists\ 1\, y \forall z [B(1,y) \vee \neg B(1,z)]$$

$$\forall\ 2\ z [B(1,1) \vee \neg B(1,z)] \vee \exists\ 2\, y \forall z [B(1,y) \vee \neg B(1,z)]$$

$$B(1,1) \vee \neg B(1,2) \vee \exists\ 2\, y \forall z [B(1,y) \vee \neg B(1,z)]$$

$$B(1,1) \vee \neg B(1,2) \vee \forall 3\, z [B(1,2) \vee \neg B(1,z)] \vee \exists\ 3\, y \forall z [B(1,y) \vee \neg B(1,z))]$$

$$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B\ 1,3) \vee \exists 3\, y \forall z [B(1,y) \vee \neg B(1,z)]$$

$$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B\ 1,3) \vee \forall\ 4\, z [B(1,3) \vee \neg B(1,z)] \vee \exists\ 4\, y\ \forall\ z [B(1,y) \vee \neg B(1,z)]$$

$$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B\ 1,3) \vee B(1,3) \vee \neg B(1,4) \vee \exists\ 4\, y\ \forall\ z [B(1,y) \vee \neg B(1,z)]$$

Part Two : Non Derivability

Mathematical proofs are very similar to games. In both there are allowed steps and we use them to achieve something. Games achieve winning situations like check-mate while mathematical derivations achieve theorems. These are special subset from a universe. In games the universe is all the possible situations like pieces on a chessboard, while in math we have the possible statements. The game starts with a non winning situation and gradually turns into one, while the mathematician is starting already with theorems that are accepted as axioms and gradually reaches the aimed statement to become a theorem too.
Also in both we have intuitive ideas that are miraculously translated into the exact steps we do.
Neither a good chess player nor a mathematician is thinking about what the next step should be.
How the intuitive process can reach the successful exact steps, is a mystery that will be now completely ignored because some much more mundane features of exactness came out in meta mathematics. Yet these mundane facts limit the derivabilities and so our miraculous abilities too.
This of course should not be interpreted as a conquest of the mundane over the mysterious!
Our age interprets it exactly this way. And this itself has a double lie in it. The mysterious becomes a taboo. An assumption that it is merely some yet unknown detail and a silent belief is spread that all things in the universe are mechanical cause and effect material processes. At the same time there is no real spreading of the simple mundane facts either. The proofs of mathematics and equations of Physics are on display but the meanings that the professionals possess about them are actually secrets. The cause is simple. Society has an agenda, a mysterious own will to keep individuals stupid. Didactics, that would spread visions to make the displayed derivations and equations plausible is regarded as blubbering by the professionals in Academia, while the real blubbering is filling other sections of society namely the Media.
A similar double lie is in Politics too. So both knowledge and actions of change are doomed.
Only individual struggle exists to see through the social lies.
This article is an example! It is a dangerous anti social activity. It spreads understanding.
Now back to facts, we have to mention some differences between games and math.
The most obvious is that the game situations are usually two dimensional while mathematical statements are one dimensional linear texts. Chess seems three dimensional but actually the pieces are irrelevant as objects and could be flat symbols placed on the two dimensional board.
The seemingly simpler one dimensionality of the texts is misleading too!
Already the concrete everyday language texts have some rules that can combine texts, but in math this has become much more sophisticated. The combinings are here done with logical operations like “and”, “or”, “imply” and so on, and are even abbreviated by symbols like $\wedge$, $\vee$, $\rightarrow$.
But these operations are just trivial special continuations.
The crucial two new discoveries of Logic were the “every” and “some” quantors.
Strangely, already Aristotle recognized these as important logical steps but he missed the crucial method to grasp them correctly. This only came to surface in the 19^{th} century because it required the use of variables that were pretty precise by this time in equations and functions. This radical step actually required that we don’t just regard the statements as universe for logic but actually all formulas that contain not only concrete objects as names but variables too.
Putting concrete names like numbers into the variables is concretization and it turns the formula into a concrete statement. But the two quantors $\forall$ and $\exists$ followed by a variable also takes away the variability of that variable and so if all variables are concretized or quantized then we also get a statement. The logical rules must be used for these too and to achieve this we needed the widening of the universe to formulas. Actually, to be really precise, we should allow any texts from the mathematical symbols and tell rules too that should be applied to even get a meaningful formula. These mathematical texts in general are usually called expressions. The so called “syntactical” correctness for an expression to be at least a formula is of course much simpler than the forming of statements from formulas. So we have three narrowing universes. The expressions, within that the formulas and within that the statements. But our present surprises will only relate to this last, the statements, though we have to step out to the others too to use the derivations.

Even in chess we have a wider universe because a toddler would put the chess pieces not necessarily in the squares and so it wouldn't be a meaningful situation. So, some pieces on a board is the wider universe and among these the correct situations in squares is the narrower.
As I said, the previous universe jump in math from formulas to statements is not our concern now but I want to emphasize the concept of variables behind it anyway:
Variables are the most important human abstraction!
The fact that in everyday languages it only appears directly by words like "everybody" or "somebody" and in games as pieces like the "joker" or "wild card" is misleading. In the thinking process both in everyday sense and in games we do use them continually under explicit language.
This abstraction is an a priori ability of the human mind exactly as Kant defined these when he restarted idealism after Newton. And the fact that we only recognized this abstraction hundred years ago historically, means again not much. Merely that we as a race are so young yet in the recognition of human intelligence. We are in fact in a long and false materialistic cycle.
But all the above mentioned details in math and games and the enthusiastic words about the future are irrelevant for the non derivability problem we approach right now.
Instead, we should start with a seemingly simple but above non mentioned operation of statements as the truly crucial difference between math and games. This is negation! Abbreviated as $\neg$A.
The axioms and rules of logic always hide the naïve assumption that for an A statement it should be impossible that we can derive both A and $\neg$A. This would be a contradiction in our system and in fact our logic is such that it would allow to derive everything. The assumption that this can not happen is called consistency. With this hidden assumption of course we can easily tell a lot of non derivable statements. Namely the opposite of any statement that we already derived.
If our axioms are not describing an intended reality, rather only provide a framework or theory of some fields that can be narrowed with more assumptions as axioms then the intention is not to be able to have a theorem from each A , $\neg$A statement pair. But the most important axiom systems are descriptive and aim at a perfect telling which one of A or $\neg$A is true in our reality.
First of all observe that even though these truths are indeed perfectly decided in a reality, this doesn't mean that we can also tell this from looking at the reality. This is so because the $\forall$ quantor means "every" and so a statement can say that something is true for all objects like all numbers. We can not check this in finite time. So deriving theorems is not just a luxury to verify the visible truths but actually to see the truths.
The big surprise turned out to be that even our intended descriptive axiom systems only work as theories or frameworks because for some A statements neither A nor $\neg$A can be theorems.
The reason why this seems so unbelievable is that for any particular such pair we could then pick one of them as new axiom and voila automatically it becomes a theorem trivially. But this just shows that our negative result actually claims that this one by one fixing up of an axiom system doesn't work and so obviously there are plenty of such open pairs that can not be eliminated one by one. This makes the previous remark about the hardness of recognizing the truths even more interesting. Indeed, we could say that lets not pick one by one new assumed truths rather pick as many as we want at once and to avoid inconsistency lets pick all the truths at once. God could do that but we can't, or rather then already our axioms wouldn't be derivable by a finite system.
So this finite derivability of the axioms is the hidden second assumption beside the first that was consistency. For most axiom systems this derivability of the axioms was trivial and thus not recognized as a condition. If the system has only finite many axioms then it seems even more trivial but actually the Logic we use to derive theorems also has infinite many axioms.
The new vision is that A , $\neg$A pairs with none of them being derivable by our axiom system is not some kind of weakness of our systems, quite oppositely a natural consequence of the strength.
So we'll shed light on some fundamental delusions we had earlier.
The basic new concept of this clarification is a natural wider meaning behind derivability.
Namely, both in its scope, that is what it refers to and in its method, so how it is established.
The new scope is not just any individual statement rather any set of statements. And the new verification is not just using our axioms and logic but any finite derivation system imaginable.
So now a set of statements being derivable means that some system that works on statements can derive exactly the set. It's crucial to emphasize that <u>exactly</u>. So the system not only must derive all

the statements of the set but only those. And indeed, the weaker meaning of deriving a set but other non wanted ones too, wouldn't mean much because deriving the universe, that is all possible objects, would make every set derivable. Though deriving a whole universe is not that trivial either but we'll show soon how this is possible.
The word derivation for a set, meaning exact derivation brings up an other word "generation" but this usually means generation in a fix order. So we'll rather call this as "listing".
In chess a listing of all correct situations from the start is quite easy by starting with a full board and then going through all the pieces on the board one by one and apply what they can do, again and again. The number of pieces might of course become less and less. And indeed that's how a chess machine is trying out the next possible situations and thus replaces derivation with listing.
In axiom systems this would be more complicated because there can be infinite many axioms.
To go in strict order we would have to order the axioms as start but if they are infinite then by trying them we never get to the next step to apply new axiom again. Plus here the axioms themselves are longer and longer so the "one by one" is stretched too. But as we see soon this can be easily solved.
While games don't have the artificial duality of negation that axiom systems have, there is a new negation concept among the derivabilities of sets that applies to games too. This is the "complement" set, the non derived situations. Then the point can be seen even better at games:
This "left out" complement of a derived set is very different from the derived one!
In chess for example each derivable situation comes about by an actual finite game history. And so we could list these by listing the longer and longer possible games.
Quite oppositely, the artificial situations that can not come about by a game could only be identified by checking all possible games and then realize that it never occurred.
Now this is not quite true because even a lame chess player like me would realize at once some impossible situations. For example, knowing that the bishop remains on its color, I would know at once that a situation is phony if both bishops of a player are on same color. But these obvious tricks are limited to special situations. The general truth is that the complement set, the whole collection of the underivable situations is a much more complex set than the derivable ones!
And now comes the point:
If a particular derivations system and thus its derived set is complex enough, then the complement becomes even more complex, namely so complex that it is not derivable by any system at all.
This is the heart of everything!
So lets return to axiom systems! What is the complement here? It is the set of the non theorems.
And if it is as complex as I claim then it is not derivable by any system.
This explains everything at once by the following heuristic argument:

If for all A , $\neg A$ pairs one could be derived as theorem then the non theorems as set could be derived by simply deriving the theorems and then negating them.
In more detail, in our axiom system the derivation is happening as a B basic derivation of the axioms and then by logic to the theorems. So $B + L$ derives the theorems.
Then $B + L + \neg$ derives exactly the non theorems if all pair would have a theorem.
But as I said the non theorems are not derivable by any system and so neither by this trio.
So the complexity of the non theorems forces some A , $\neg A$ pairs to stay in limbo.

The problem with this heuristic argument is twofold:
Firstly, we didn't specify all possible derivation systems and so an impossibility of any such has no exact meaning. Secondly, we didn't give exact conditions when an axiomatic derivation system is complex enough to make the complement set, the non theorems over complex.
So, we should go in these two directions. Avoid the undefined concept of derivation systems in general and find exact conditions of complex systems. Still, the essence is above and all others are background details.
A particularly useful detour is the comparison of the sets of theorems or reachable game situations with the sets of derivations. The derivations are just finite sequences of theorems or consecutive game situations. Though this is not quite true. These sequences must contain the claimed

derivational steps in-between as applied rules. These are the allowed moves in games or the usable rules in logic. So the universe of derivations has a bit wider language than the situations or the statements. Namely, the rules must be named too and we then regard sequences with these as the connectors. Just as any set of theorems or reachable game situations have their complements, derivations also do and these are the false derivational sequences. It can be false by false chain pieces that is situations or statements, or by false connections that is false or not applicable moves or rules. But here this formal falsity at once reveals the complement elements. Indeed, such crazy sequence or faulty derivation is very easy to recognize. We simply go through and check what rules were employed at every step. If we are dealing with a non valid derivation then after finite steps we'll spot an error in our chain. That‘s why old proofs can not have errors. Because simply so many mathematicians went through them already. Also that's why chess championships can not have cheatings. Too many people are watching. What all this means is that unlike the sets of theorems and reachable game situations that I claim to be such that their complements are non derivable, among the derivation sets there is no such asymmetry. The correct and incorrect sets of derivations are both easily recognizable and this implies that they are both derivable sets too.

Dual recognizability implies dual derivability.

Instead of making legal but freely chosen derivation steps we start with a freely chosen object.
This was any claimed derivation sequence in our example. We do the recognition steps. To see that all derivation steps were correct in our example. This takes only finite time to finish.
To derive the opposite set we do the same for the complement dual set. In our example this means to look for error and stop at once one is found.

The same is true for a stronger version of derivation that avoids choices and goes in fix order and we called it as listing:

Dual recognizability implies dual listability.

The basic idea is very simple. We list all possible objects but not claim this as final list rather do our two complementing recognition processes simultaneously for every object one by one.
We never get stuck because one of our recognizers must recognize and that tells also which list we continue with that recognized object.
The only problem is the start, that is to list all possible objects. These were derivations, that is sequences of statements or game situations in our example. Already the statements and the game situations themselves are complex objects not mere numbers that can be trivially listed in increasing order.
So involving all connecting steps into an alphabet, at the end the task is simple:
How to list all possible texts from an alphabet?
Our human dictionaries going letter by letter only work because deal with limited words. Here at texts we have infinite many texts starting with A so we could never reach letter B already.
The trick is beautiful: We have to go in increasing lengths of the texts. The one letter long texts are simply the listing of our finite m letter alphabet. Then the two letter long ones will have m^2 many and so on.

Our result is true for even non dual sets:

Recognizability implies listability.

The universe set again can be listed by our smart length increasing order but now without dual recognizings we will get stuck. Indeed, the non recognizable members in our list are obstacles that we can not jump through. We could wait for ever to see that we have to step to the next object.
We need a second trick! We will start our recognizing process on all listed object anyway.

Of course we can not start on all infinite many so we start on the first. We do a few steps of our process and stop. Start processing the second object on the list and again stop after few steps. We do this for a few and thus get ahead. Then we return to the start and continue our processes for all we started. Then we go even further to start new ones but again return to the start. As we see, we can reach more and more objects to examine for longer and longer. Thus eventually all recognizable ones will finish and can be produced as next in a list. Of course this list will not follow the original list order because quite far away ones may pop up as recognized before much earlier ones. Still, since we do this for ever, all finite recognitions will come about.
This method is called dovetailing.

The reverse of our result is quite trivial:

Listability implies recognizability.

Indeed, now all we have to do is wait! When the object to be recognized pops up in our list we claim it to be recognized.

Of course, all these were totally irrelevant for our basic enlightening that the underivability of the non theorems causes the existence of undecidable statements but will be useful later.
So, we have to go forward in the two directions. Avoid the vague concept of undefined derivation systems and find exact conditions for concrete systems to be complex.
In Gödel's original article the heuristic idea was not recognized! He had to sweat out a "concrete" pair of undecidable statements and himself regarded the language of the axiom system as a main cause. Soon however the truth was realized by two people. Church and Turing. Both used a common smart way to make their points and go beyond merely claiming the above heuristic argument as some philosophical claim. The crucial new idea was to go in-between the undefined loose notion of systems in general or just regard concrete known systems. This in-between is then to define a particular whole set of systems called as a "framework".
This beautifully addresses both problems.
Indeed, now the vague concept of systems in general boils down to whether the framework has members that can imitate any imaginable system. So this is an external imitation and universality of the framework as whole. This still can not be derived precisely but is much more concrete at least. Also observe that it could be that the framework is so grand that any imaginable system fits into by definition at once. So then anybody's suggested system would be visibly in the framework by simply showing how the definition can imitate it. Unfortunately, such framework still doesn't exist and imitation of challengers is thus not obvious and boils down to show that imitation exists in the sense of deriving the same objects. Turing's framework became the winner because it had a vision that slowly started to evolve into almost a plausibility for external universality.
The reason of this is that Turing had the future computers in his mind which means the concept of machines as such too. Machines not as matter working on other matter, rather as operating by strict rules. And it seems that we have some a priori intuitions about this concept. Of course I would argue that games for example don't fit into this and we still have intuitions about those too. So there might be some wider and better base of intuitions in the future.
But back to our claim that their framework idea addressed both problems, the really important breakthrough was that they found out the complexity condition inside the framework.
This is a new internal universality, meaning those particular systems that can imitate all others.
Of course the word "imitate" is still an imprecision but as we'll see it can be made very precise though unfortunately again, it never became plausible. And yet, it's interesting that in this respect Turing's framework again became a winner. Indeed, at modern computers this internal universality is nowadays claimed as "plausible". But this is a false claim. We "feel" convinced that different operating systems could imitate each other but only through our familiarity with the details and not by a plausibility of the framework as computers. Simply, because this framework of computers in general is actually quite foggy in spite of everybody owning a computer.

So, Turing realized that computers are the future. And it became the necessary present for him to create a framework where this internal universality can be demonstrated. But this had to be well defined not foggy like computers in general. So, preferably as simple as possible, yet be able to do everything that real computers of the future could do. So he had the future flexible computers in mind but made a very non flexible machine framework to use.

The crucial complicatedness of real computers is that they work parallel at many levels. The electrical signals go many different places and do many things at the same time. A consequence of this is also that spatially the different parts where data is stored have to be accessible. They have to have addresses. Amazingly, this is avoidable! We can dump everything on one single line after each other and let the central processor unit, the famous CPU from the film Terminator, find every detail. So what we have is a "crawling" computer. Every single bit of information sits on a single line and every time we want to deal with some segment, we have to crawl there and bring them bit by bit away to a free part of the memory line and do there the calculation. If we need the result later then we even have to take them back to the memory section of the line. So, though Turing's computer could do everything that a modern computer does, the simplest tasks would take millions of years to execute. But who cares? We are in mathematics, time is irrelevant.

But something else is very relevant. Infinite space! In spite of that we dumped everything on a line we have something incredible in Turing's computer that real computers don't have. Infinite memory! To spread a real computer's bits to a single line would require billions of kilometers but now again we can say who cares. There is always room for more on an infinite line.

So this crawling computer is the first half of the framework that Turing used instead of derivation systems but the actual representation of a particular system as a particular machine is not revealed yet. This first part was based on his belief in computers of the future and that his crawling one can imitate the more practical ones of the future. To see that this limitation of a single memory line and cell by cell alteration allows everything is pretty soon becomes obvious to anyone who tries to convert some goals into such crawling jobs. But his second vision of the particular machines as simple "tables" is not intuitively obvious at all. Even if we play with them a lot. So I will come to that in more detail after I show his original ingenious idea:

As I said the CPU goes cell by cell and obviously has to be able to move in both directions.

Plus it rewrites any cell, which of course includes leaving it unchanged by writing the same as it was. But how should this decision of rewriting and move left or right be made? And by what?

We could not merely make the rewriting decision by what we see at a single cell. So one would then get entangled in a new computer to be built in to store earlier read cells. Amazingly such CPU memory is not necessary, or rather it can be replaced by so called "states" of the CPU.

The state in which the CPU is at a moment is the crucial additional deciding factor for what to do. What to rewrite from our alphabet, which way to move and which new state to enter.

Just as the alphabet is the possible external choices that the memory cells can contain on the line, the states are an internal alphabet that the machine itself takes up. Or they could be called as the "moods" of the machine.

The actual operation of the machine is determined by the transition table that contains for every possible read symbol and state pair a triplet of the new symbol, left or right motion and new state:

(symbol , state) $\rightarrow$ (symbol , left or right , state).

If we have m many possible symbols and n many possible states then we have $m\,n$ many such lines in our table.

Starting from a line of our table and from any cell of a fully loaded memory line we enter into an infinite process of jumping to new lines in our table and rewriting the line cell by cell.

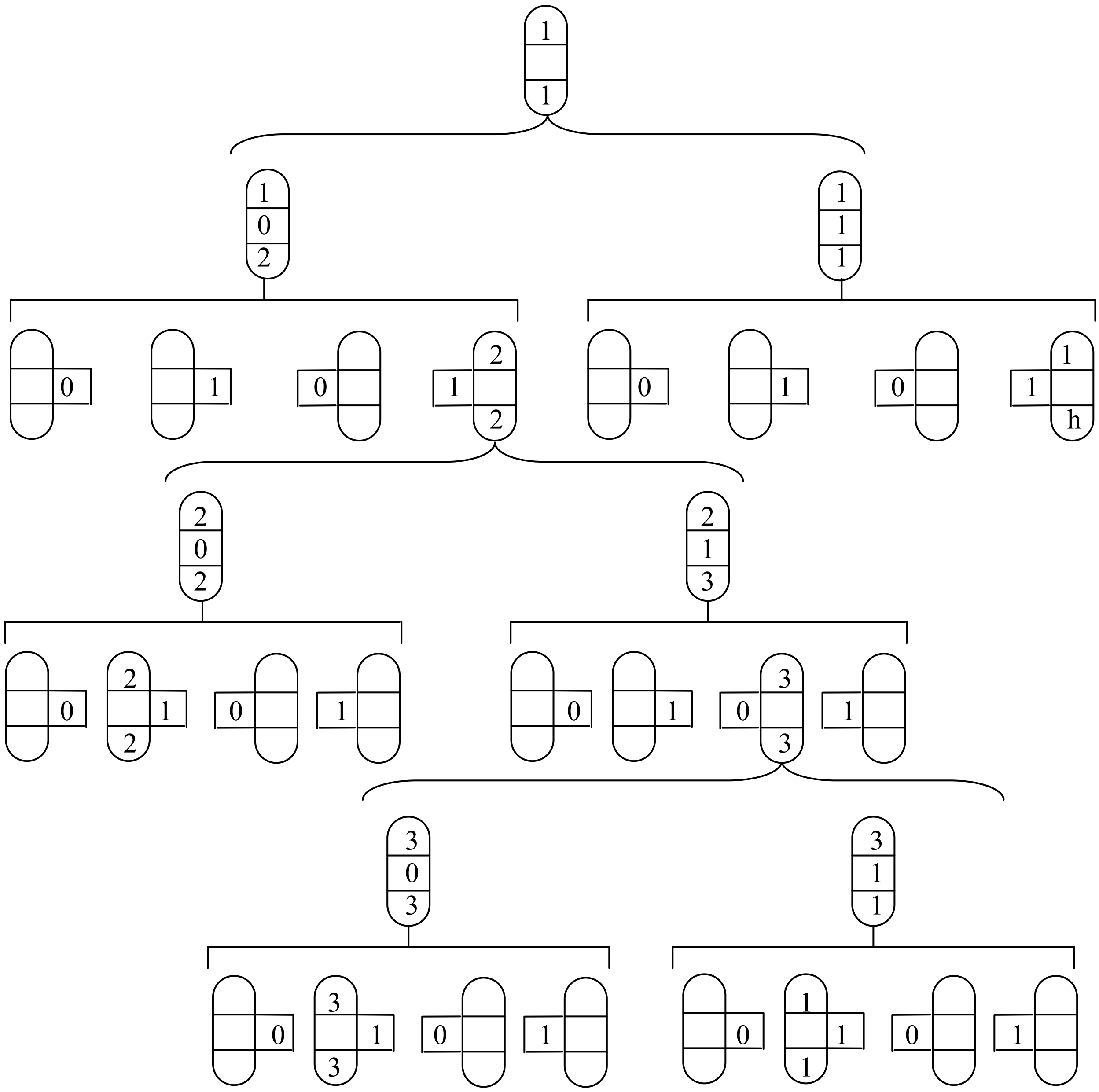
The big difference is that the table is finite but the line outside is infinite. That's why the table jumps won't become some boring cyclic process rather can become always surprising. Some states may only be encountered by arbitrary many jumps because the external conditions on the memory line influence the internal jumps. This explains why we can even choose a starting and halt state and still have flexible jump sequences. Then these of course can only be two kind.

Terminating if we reach the halt state or non terminating if we don't.

The memory line alteration is determined by the table and the chosen start and halt states but the initial line content if were arbitrary would mean an infinite amount of information. So this could only be called as a relatively effective method. Proper effectivity would require that the initial memory line is trivially effective and simplest such is having all same symbols called as “empty”. A crucial added flexibility is allowing a finite non empty initial segment on the line as “input”.

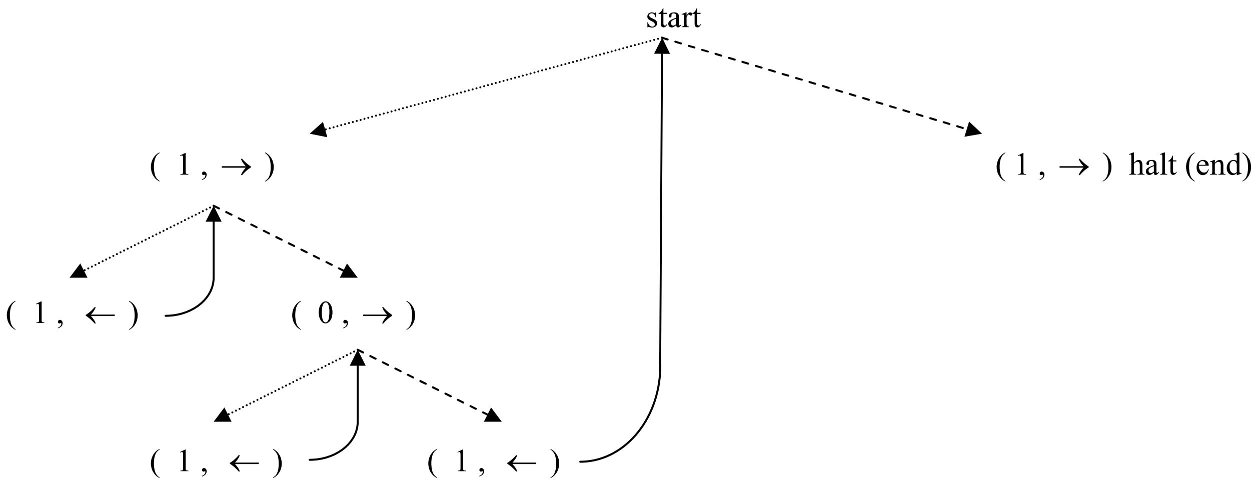
The big question is how could we be sure that the table method can bring about all possible effective alterations of the line. We can’t. This table vision was a truly lucky guess from Turing after playing with a lot of concrete machine goals.
But there is a way to make this table idea a lot more convincing if we start from the first part, that is the crawling idea on its own. How could the CPU go if we wouldn’t have the table?
Well randomly, or as it makes up its own mind about what to rewrite and which direction to move. So $2m$ possible choices can be made at each action and to be at least meticulous we could make an action tree that consists all possible choices for the CPU. The CPU will be pictured as a bean shaped unit that scans in its middle the cell and we’ll draw that symbol there. This could be m many drawn as going from “a” to “z” but of course including all symbols including the blank too with a definite symbol because this way the empty middle in the CPU will not mean blank rather that it hasn’t been scanned yet. This intermediate state is useful because underneath then we’ll always have a round clumper containing the m possible read symbols and under each of these a squared clumper will contain the possible actions depicted as the bean again being empty but next to it showing the written symbol. If it’s on the right then the CPU moved to the left if on the left then it moved to right. So we see it very visually what was the action. And again under in round clumper we reveal the possible symbols. The only exception is the first and second line where there was no action yet so we’ll have no neighboring cell shown.
This infinite tree with all possibilities is the roadmap where all actual infinite action sequence could happen. Such actual sequence is simply a path down picking new members from the continuing round clumpered symbols or squared clumpered actions. Observe that the picking is different for the two clumpers though. The round ones are actually determined by the line symbols and so if the original line content was known then we can see these though not all will be original because our earlier rewritings may be there instead. The square clumpers on the other hand are totally our choices indeed. So a path would involve these infinite many square clumper choices. Of course infinite many random choice is not a machine and so the task is to “prune“ our tree so it should have only finite paths. But then how could we make infinite many actions? Simply redirecting our prunings to earlier branchings and from then on continuing according those earlier choices in the square clumpers. Now comes the crucial question. If I want to prune a branching does it really mean to cut it right there? No, I can just go on and cut all the continuations at different points. This of course would make things very complicated. Indeed, then the return point for each continuing point would have to contain the continuing localization for it. Turing’s state idea is equivalent to the assumption that simple local prunings can achieve all overcomplicated gradual prunings. But still even with this simplification, to give the square bracket choices seems like a mess. To give a number to all possibilities and choose one above in the beans lower section would be the logical thing to do but since we don’t care about the choices not made, we should simply use only one number in each square bracket, namely in the upper place of the bean where the chosen action is. The action itself will always keep this same number in its bottom too because it can only be split underneath in the round clumper. In the round clumpers then the lower numbers can now be altered again from the top, namely telling where to go in the square clumper. Now comes a great trick by allowing these single used numbers in the square clumpers to be an earlier. This combines the two tasks, the specification of the actions and the prunings as well. Indeed, if this number is an earlier then we still go into the square clumper to the marked action but then we don’t open new round clumper under just return to the earlier.
The n number of used round clumpers are exactly the states and they can be numbered increasingly as they appear. The number of used actions are mn many because under each symbol choice in the round clumper we shall give one chosen action. And indeed these are the possible lines of our table. The read symbol is above the square clumper in the bean, the present state is there too underneath and also in the square clumper action repeated above, while the action is visible in the cell next and the new state underneath.

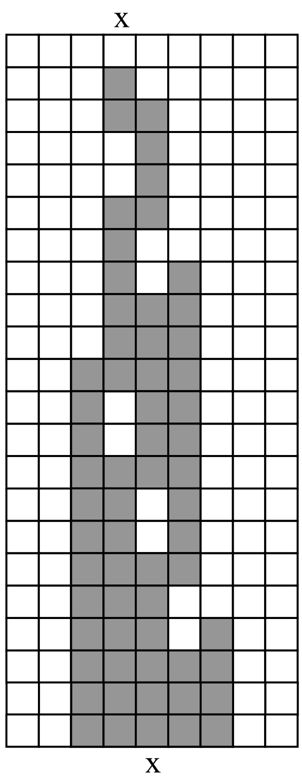
This tree representation has a definite start and thus it shows the starting state as 1. A similar usable convenience is to do a drastic pruning without any return so here we use a last state denoted with the letter h standing for halt or termination. So we don't draw round clumper from this with possible read symbol choices because we are not interested in those.
Usually this halt state is not counted among the normal numbered states.
As an example of a pruning, I give a famous case, the so called two symbol three state "step busy beaver". The two symbols are 0 and 1.



A major simplification can be achieved by avoiding the square clumpers and use only the chosen actions in them. Then as a further step we can even replace the round clumpers with possible arrows that show the read symbol. For two symbols 0 and 1 a dotted arrow can be for reading 0 and dashed arrow for reading 1. The returns then can be marked with solid arrows back to an already defined read symbol split. So, here we just go down again. Thus the states completely disappear and we get a simple "action web" that tells the most visually what really happens on the line.

To prove my point, I will show the action web and right under the actual line alteration starting from an all 0 line, using empty squares for 0 and then grey ones for 1.
The two x show the CPU at the start and at the end.





As we see, we made 21 moves and left five cells as 1.
With a given number of states, that is clumpers, that is arrow junctions and one halt, we can only create a certain number of such pruned trees, that is tables or Turing machines or action webs. This is quite easy to calculate too, how many.
If we look through these and check one by one how they behave then the major two kind is whether the halt is encountered or not. Those that do get to halt will have a certain number of steps and written 1-s. Tibor Rado asked the question how many can these be. Obviously the 1-s can not be more than the steps and the amazing result is that even these maximum 1-s will exceed the number of possible machines astronomically. In fact, they will exceed all calculable functions. Then of course both these maximal 1-s and the maximal steps are not calculable.

Originally, those machines that produce the maximal number of 1-s were named as busy beavers and later the ones that produce the maximal steps became the step busy beavers. The example above is such and so with two symbols and three states there can be no more than 21 steps if we halt at all. The left 1-s however can be more, namely six with a machine using only 13 steps.
Knowing the step busy beaver of course is much better because then we know how many steps to wait at all for a halt and thus we can check out with a computer all possible machines only up to this many steps and will see the maximal 1-s and thus get the busy beaver too.
Already for 5 state we don't even know the busy beavers yet. We only have values that the 1-s and steps must definitely exceed because some concrete machines already produced them.

All tables are finite but the possible tables are infinite many because we allow arbitrary large number of states and thus tables too. And that is what makes the already foretold internal universality sound so surprising first: There is a single finite table that can imitate all others!
The trick is to use the memory line but also the word "imitate" must be interpreted flexibly.
Perfect imitation, that is doing exactly the same elementary crawlings is impossible. The only thing we can expect is that some part of the line regarded as a program will be used by the finite universal table to alter the rest of the line exactly as if this program section were not there. Since this program section can be arbitrary long thus this finite universal table imitating all arbitrary big tables is not a mystery anymore. The imitated table is "simply" stored in the program line by line.
I used the quotation mark because this is not a simple affair. A linearization itself merely would need markers to separate the lines but lets not forget what we have to do with this program.
We have to use it to alter the rest of the line as the table would that is stored in the program.
There is basic problem as start! If we want to imitate arbitrary large tables, we have to imitate arbitrary many states. To introduce a symbol for all in our program would mean that we can not have a fix alphabet for our programs. The solution still lies in the simple fact that the program itself can be arbitrary long. Indeed, we don't need different symbol for all states because we can use a single symbol repeated and its repetition would tell the state. In fact, we can avoid states too by using the above shown clumper system which is actually a web of actions. But to identify the actions we still need the repetition idea. This idea hides a further detail, namely how to locate the action by these repeated addresses. In short how to count these repetitions. This is to use cursors. These program symbols can replace the repeated ones and move forward thus telling if we went through or not yet. The actions themselves can be replaced by single program symbols because their number is finite for all machines with a fix alphabet. For example with two symbol alphabet we have four actions. Writing one of them and moving left or right. This is what we have to deliver to the data. This raises two problems: How to deliver it and where exactly. The second seems easier by a data cursor while the second reveals the exact detail of what I said earlier about the states replacing a CPU memory. We can not put one of the four possible actions in our pocket but we can create four alternate states for all normal states of our program. So the CPU reading a particular action symbol in the program will shift its state into this symbol variant and thus "remember" what the action was. When it gets to the data cursor then the state will tell exactly the action or rather trigger the right action by the read cursor. The real beauty is that the data cursor doesn't have to be inserted into the data. It can always replace the next read symbol because the same way, that can be stored too as alternate state that the CPU delivers back to the program. Actually, for a moment when the CPU writes the new symbol and steps next, the cursor symbol disappears. It will only appear after the that next symbol is read and is replaced by the cursor symbol. If these snapshots were played after each other and the program section also cut out from the line, then we could see exactly what the table would do that we imitate. So, all those longer time intervals while the cursor is intruding the data, searches the program and comes back would be cut out in time. And in space too the section where the program is intruding the line.
An interesting situation is if the data line is supposed to be altered jumping through the program. This is apparent because the data cursor is ordered to go into the program prematurely. So it's easy to make this correct by the program and continue on the other end.
In order to use this grand imitation idea for our purposes we have to make some detours:
An important issue that was handled already but very superficially is the halt state.

This is a deep stuff! In fact, it is the third heuristic trick from Turing beside the infinite memory line and the states. Indeed, normally we would use our alterations until the data on the line reaches something that we aimed for. So the termination would require a new computation to decide whether we achieved what we want or not. But just as the CPU memory was avoided this can also be shoved into the states. The fact that they can cope with this, is far from obvious though. This also shows how non trivial the external universality of Turing machines is.
In fact, their use as derivation systems is not even apparent after we revealed this crucial termination affair. After all, if the altered data is used as situation checking then it might not be the actual derived result. And indeed, usually this is how we use a Turing machine. Not regarding the alterations as striving for a "result" rather as striving for the recognition of an "input". The recognition is when termination happens. And derivation is only consequential as we explained for all recognizers. This conquest of the recognition concept over the derivation, as practical detail turned out to be true even for the logical derivations too. Namely, the original variable logic of Hilbert can be replaced by a much nicer system of Tait. This can be regarded as a simple derivation system too starting from formulas, but actually we can start with any statement and go backwards too. So it recognizes totally mechanically all statements that are logically derivable. Then an additional side result is that when this recognition process is not successful because it goes on for ever, that is without a halt, then this infinite process creates a model where the statement is false. So the system is actually proving its own completeness instantly!

The input idea would suggest that we simply have to mark the beginning and end of what we regard as input. But this is not true because the machine must be allowed to go out and do whatever it wants to be complex enough. So then outside this marked segment a lot of hidden input were still present. To avoid this, we'll distinguish the blank symbol and then the data is the single segment without blanks in it but between the infinite blanks on its two side. The machine can again do whatever it wants but there is no hidden information outside the input. It is allowed to use the blanks as we use an empty paper. The single segment that was our input will scatter all over on the line as our calculations on a paper. This shows already why the recognition use of the machine is better than an actual derivational use by the created alterations. Indeed, those would be scattered pieces too. Our goal is to show that derivable sets exist with non derivable complement. And this way the universe is merely these possible finite input segments on the blank line.
A surprising special case is if the alphabet is the minimal with two symbols. We would think that by the specialty of the blank, this would mean two other symbols. But this is not so! One is enough! So a single 1 symbol repeated is already a possible input. Indeed, these are the natural numbers as inputs. The monotony is misleading. A certain number of 1-s hides many secrets and can be explored by alterations. Scattering it for example we can translate it into number systems with using the blank as separator of the digits. We can check dividabilities, primeness, everything.

Before the crucial trick we should introduce two very smart concepts that avoid a lot of rubbish that has been said about this final trick.
Let's regard all tables with a fix alphabet including the blank and all containing halt state.
The blank is excluded from inputs and:

We call two tables "sharing" a d input if either both halt from d or both do not.
We call two tables "complementing" if one halts exactly at those d-s where the other doesn't.

Evidently, the two concepts are the exact opposites:
Two tables share an input if and only if they are not complementing ones.
Most importantly also evidently:
An S set being recognizable by a T but $\neg$S not being recognizable by any table means simply that T has no complementing table.
So, to prove the existence of such S it's enough to find a T that shares input with all tables.
The idea is this! We look at the given alphabet and create our special alphabet for our universal table. This will be definitely bigger because we create symbols for all actions.

Now we want to replace this bigger alphabet with the simpler that our chosen set of tables use for inputs only, that is without blanks.
This is easy by using sufficient many symbols to abbreviate a single program symbol.

For example, by using groups of fives from merely two symbols, we can abbreviate $2^5 = 32$ many new symbols already. Of course, with the mentioned single 1 symbol inputs, the situation is problematic! Even the mentioned blank separated digital forms are still to narrow.
So we assume at least two data symbols and we start with any such input.
First copy it next to it exactly after a blank. This is strictly doable by a table from our set.
We continue with the process of translating the left one by groups of five symbols into a program symbol text. We might have an input that its length is not dividable by five and so this coding fails already but even if it is okay, we can get a totally meaningless text that could not function as a program. But this doesn't matter! If we do get a proper program lets start to run it on the other version of the input next to it.
For every T table there is some proper program for it and so there will be an input that codes into this program. Now lets see how such input would be altered by T. We have no clue. The table has no clue either that this seemingly random input codes into its own program. But one thing will be sure! That T will share this input with our process. Indeed, our process started with the same input and after duplication it will start to imitate T with this very same input.
So, a halt of T is exactly a halt of ours while no halt of T means that ours won't either.
The six million dollar question is of course how such alphabet change could be imitated by table.
To run the program itself by codes of fives is not a major difficulty but to actually go into the data with a five symbol cursor is insane. Not mainly because the five as group must move together but because these symbols in the five are normal data symbols and will melt into the data.
The solution is drastic! We have to split the whole line into alternate even an odd cells. The data can only be in one, while the data cursor in the other. So we actually have two lines combed into each other. The blanks in the cursor line can be distinguished from the real blanks in the data line and so we can still use these as before. Of course, the moves are now double steps. The big advantage is that now we can use any of the data symbols as an uncoded data cursor that directly shows where we are on the other data line and yet can be distinguished as not data.
The program can also be split and put into the non data line just to make the system cleaner and also avoid the jumping through it. This was how Turing used the line too from the start.
Of course, this method needs now at least three data symbols. One uncoded as data cursor and two others to code the program symbols. After duplication, we choose one as data cursor and create imaginary program symbols from the others as groups. But then we actually split both copies. The program is of course altered to use this. The most important consequence of this drastic change is that now the exact imitation of a T as alteration of the original data disappears.
But the imitation of the termination remains. So the input sharing remains too.

Now that we demonstrated a table without complementing one and thus also that the recognized or derived input set of this is without recognizable, that is derivable complement, we have to ask, did Turing really reproved what Gödel did in his original article. The existence of undecidable statements in number theory. The answer is no! Turing's article is quite an opposite of Gödel's.
Clear expression of the new spirit that non derivability of sets is the cause behind the undecidable statements, but pretty messy details. The existence of universal machine is already pretty messy.
And then the consequence as non derivable complement is not as clear as by our "sharing" argument either. But his proof is exact in the details. Then Turing only sketched how the number theoretical derivations can be simulated on his computer. Luckily, the details of these are totally mechanical. This however is still not enough.. We have to see why the set of theorems in number theory would also be such derivable set without derivable complement. To achieve this, one would have to show that Turing machines are also representable in number theory. Then the Turing created set with non derivable complement were still only a particular set of statements with the non theorems being non Turing derivable, but this would easily show the same for all theorems and non theorems because these particular formed statements are easily recognizable by

a Turing machine. So then the heuristic end argument would show negation as a trivial step and so the full decidability would make it impossible that the non theorems are not Turing derivable.
There is a very smart way of seeing when a proof is imperfect in details. We show a case where the claim is false yet the sloppy proof would still “prove” it, which of course is impossible. And here we have a perfect example for this! Gödel used many number theoretical details in his proof about dividabilities. This was so because he regarded the language of addition and multiplication. Now if we omit multiplication and only create a baby arithmetic for addition then it is possible to have total decidability, though to prove even this is quite a task. Pressburger showed it a year before Gödel proved that with multiplication things turn nasty. Turing’s article can give the impression that the Pressburger arithmetic should have undecidable statements too. Simply because some crucial elements of Gödel’s argument are missing. Namely, why things turn nasty with multiplication. The simplest answer to this is: Because multiplication has the power to code all finite sequences of numbers with a single c number.
More importantly, in reverse too for any single c number we could create the i-th element of the c coded number sequence with a single T (c , i) formula function.
All this is due to the fact that dividabilities or remainders of huge numbers can produce arbitrary long arbitrary membered sequences.
The shocking consequences of this T (c , i) sequence generator can be best illustrated by exponentiation.

$x^y = z$ could only be defined as a new case by case obtainable relation through multiplication that is added as basic relation to the axiom system. Or we can be sloppy and do it as in elementary school. That is, allow dots and say that it is simply repeated multiplication:

$x \bullet x \bullet x \bullet \ldots \bullet x = z$ with y being the number of appearing x-es.

This is not a fix formula so it is not a strictly inside definition and we might even feel that such fix formula would be impossible. But with T (c , i) it is child’s play:

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

Here ∃ means that there is such and ∀ means that for all.
If there is such c code that for all i under y the bracketed three claims are true then the first tells that the sequence starts as x the second that the members always multiply by x and finally the third that the last member is z. So three claim could express the dotted non fix long one.
Now we can see that this at once changes everything and all similar rule definitions can become finite formulas on the language of a multiplicative arithmetic.
For Gödel his proof showed that the language of multiplication is strong enough to talk about the system itself and this caused non decidable statements. This was a false view but still alive!
The truth is that multiplication is strong enough to make the non theorems to become a non derivable set and this forces to have undecidable statements. At the same time multiplication creates formulas that we regard as self descriptive and through these we can catch concrete cases of undecidable statements. Unfortunately, if we try to use a fix outside derivability system like Turing machines to replace the imprecise concept of derivability as such, then we have to show that that external system is representable in the axiom system. This will require the sequence codings and so we are back to square one. We have to prove not less but more than Gödel did.
And so indeed, seemingly the caught cases of undecidable objects will dominate. Luckily though the bigger picture still shines through our distorted subjectivity that chases proofs. Simply by chasing wider systems where undecidability has to be. We’ll come to these at once.
But I have to return to Turing’s article with an other criticism. As I said the internal universality was fully demonstrated though avoiding to dig into the nasty problem of the common alphabet for the two inputs. The external universality, that all derivabilities can be simulated by his machines is of course not provable exactly because we do not have a strict meaning for all possible derivation

systems. Just think about how many new games could be constructed. And this is mentioned by Turing too as a subjective non strict assumption. Yet he still claimed a weird external universality.
To see why, we have to go back in time when Hilbert collected his famous list of problems to be solved. There were two among these that were not fix statements to be proved, that is hypothetical theorems, rather to find some method that would decide two lists of problems. The first was to find a method to select the purely logically derivable truths in the universe of statements. These are the ones that are derivable by logic alone without any axioms and therefore are universally true in all realities. The second was to select those number theoretical equations using elementary operations that have solutions. Observe that both tasks have one side that are easily derivable as sets. Indeed, the logically true statements can be derived by listing all possible logical derivations for all possible languages. The infinity of possible languages is not a real problem because we can use our already mentioned trick and not go in strict order. So instead of sticking to one language, we can list all possible more and more complicated languages. Then start deriving the truth in the first then stop go to next language, stop again go to a few but always go back to the start. Go in dovetailing. Quite similarly in the other problem, the solvable equations can be also listed by searching for solutions one by one. So the real problem is that the other halves, the non logically derivable statements or the non solvable equations are not derivable at all by any system. If they were, we would have selectability trivially.
But all this is merely "according to our present knowledge".
Because non derivability "by any system" is not a fix mathematical claim. Indeed, we do not have a trivially universal concept of derivation systems or rule systems or machines.
I said this but others would say that these Hilbert's problems were proved negatively.
Turing also claimed that he disproved Hilbert's quest for a method to tell all logical truths.
But of course exactly his other words in the article show that he new it is impossible.
Turing merely made it very probable that Hilbert's quest will end without solution.
This is so because Turing showed that his machines can not do it (though even this had gaps) and because he also showed how his narrow crawling machines can be used to do all computations.
Now if someone would claim that he found a method to select the logically true statements and the rest then he would have to not just prove that indeed his method correctly selects the statements into the logically derivables and the rest but that it is a method deciding in finite time.
I argued above that it also means derivation of both sets but this is irrelevant. There is no fix set of all possible derivators nor recognizers nor selectors or any mechanical systems.
As I already mentioned, in the same year 1936 as Turing's article appeared, Church published one where a much more abstract derivation method was introduced than machines. In this "lambda calculus" again the internal universality was the exact part implying a non derivable complement.
So the same grand vision of non derivability of sets being the cause of undecidability of statements was realized as by Turing. What's more, the same reluctant struggle is present in the article to claim the framework as externally universal too. Later this was even named as Church Thesis and the similar for Turing's assumption as Turing Thesis. Nowadays combined as Church Turing Thesis it just means that the set of all effective systems is an entity outside math at present that we can still exhaust by these two concrete frameworks. In other words, whatever text sets any hypothetical system could derive, some system of these frameworks can derive too.
Gödel's reaction was strange. He at once rejected the Church thesis as intuitive and worked on his own framework as general recursive functions though never was happy with that either. Later he admitted that Turing's framework as machines is superior to all others and nowadays this is taken out of context to elevate Turing's importance. Which is stupid because his genius is clear anyway.
But the really strange "end of the story" is this:
Kleene in his "bible" called "Introduction To Meta Mathematics" introduced his own framework and it is a continuation of the old primitive recursive functions too but not as Gödel did by intuitively widen it to everything, rather oppositely to allow one single new step. So this is indeed the most concrete and precise framework for deriving number sets.
In his book he meticulously proves theorems that allow to derive the grand expectations from a truly externally universal system. But these actual arguments are always machine visioned ones and so in all new textbooks a crazy triality exists. Kleene's partial recursive functions as the spine,

Turing machines mentioned as alternatives, and then in the actual arguments using just all possible machines, so accepting the Turing Thesis. All this is a necessity and so the machine plausibility of humans is an unavoidable fact that shouldn't be a taboo for the sake of being exact.
Kleene tried to make his system to be more plausibly universal and in 1981 he presented a version that he claimed to be such. But this ignores what I just told about the split between arguments as plausible flows and exact verifications as proofs.
The split between arguments and proofs hides the mystery of plausibility as the real skeleton of understanding. This is ignored by everyone today. Only a future Didactical Logic will resolve this.
Now we come to the promised wider results that show that underivability of the non theorems is the real cause of undecidable statements.
The big surprise turned out to be how little is enough so that a system could talk about its own derivability. Gödel regarded the full strength of number theory even though he actually used a tiny fraction of it in his proof. A genius like him should have realized this, so why didn't he throw out the unused assumptions? Because in this particular case the more was needed to make his point not just to prove his point. Indeed, the point was that even this strong number theory must have undecidable statements. If he threw out the unused section of number theory then it would have contained undecidable statement right at the start trivially. So then his proof would have been superficial. So this situation shows perfectly the modern view that non derivability of sets is the reality that creates undecidable statements.
The mentioned portions of number theory are also very simple.
The seven most obvious axioms define counting x' , addition x + y and multiplication x • y.
Counting is the real ground, it is not based on other like addition on counting and multiplication on addition. This also means that quite surprisingly, addition and multiplication is much better described than counting. This sounds unbelievable because counting is so primitive.
The fact that every number has a next and it is unique we don't even have to say because that's assumed for any function and so the next function x' obeys these too.

The first fact we claim is that 1 is not next of any number.
The second is that everything else than 1 has a previous.
The third is that the previousness is unique:
So if two x' and y' numbers are the same then x and y are too.

The 1' is what we abbreviate as 2 and then similarly (1')' = 1'' = 2' = 3 and so on.
So the concrete numbers as names are just special texts too, with many repeated next symbol.
We might think that these three axioms exclude everything else than these normal numbers, but we are wrong. Adding a single extra ω "meta number" to our naturals of course indeed would be impossible because every number has to have next by our basic function. But adding ω' and then ω'' and so on would still not be enough because our second axiom says that only 1 has no previous. So, we have to have previous of ω which could be denoted as ω_1 and then the previous to that as ω_2 and so on. This . . . ω_2 , ω_1 , ω , ω' , ω'' . . . double infinite snake could be the mysterious foreign objects of our axiom system. Strangely as we go further with more and more axioms this only becomes more and more complex but can never be avoided completely.
Addition has two axioms. One tells that it is just the next function if we add 1 :

x + 1 = x' The other tells the next value if y is increased by 1 :

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

Amazingly, this will tell all sums. For example to derive 4 + 3 = 7 :
4 + 1 = 4' = 5
4 + 2 = 4 + 1' = 4 + (1 + 1) = (4 + 1) + 1 = 5' = 6
4 + 3 = 4 + 2' = 4 + (2 + 1) = (4 + 2) + 1 = 6' = 7

Finally, multiplication again has similar two axioms to be derivable from addition.
The first tells that multiplying with 1 is just itself :

$x \bullet 1 = x$ The second axiom gives the value for a next y with addition :

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

As unbelievable as it sounds, these seven axiom could have been enough for Gödel's proof.

But this system of Robinson can obviously not decide all statements. In fact, we can not even derive the simple fact that addition is exchangeable that is $x + y = y + x$.
So as we see, just because we perfectly derived the arithmetical operations, it doesn't mean we can tell the truths about them too. To do that, we need a new derivation of not the objects that is numbers rather the properties about them. The idea of this is what we call the induction axioms.
This claims a P (x) property to be true for all x values if it is true for 1 and also inherits from x to $x' = x + 1$. There are two incredible hurdle with this.
Firstly this has to be claimed for every possible property, in fact for all formulas with one variable regarded as x here and the others staying fix. So we need infinite many axioms.
Secondly, for claims about more variables this inheritance process has to be carried out for all variables one by one. This makes the derivations quite complicated.
For example the $x + y = y + x$ feature is about two variables and so we first use inheritance to show that the $x + y = y + x$ exchangeability is true for $x = 1$.
That is : $1 + y = y + 1$. The $y = 1$ case is trivial: $1 + 1 = 1 + 1$.
The jump from y to $(y +1)$ can be obtained as :

$$1 + y = y + 1 \quad \rightarrow \quad 1 + (y + 1) = (1 + y) + 1 = (y + 1) + 1.$$

The used tricks are the two equalities in the consequence to get the ends as result.
The first is our addition rule with $x = 1$. The second used the assumption itself.
Next we show that a first member incrementing rule is also true for addition:

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

For this, the $y = 1$ case is trivial again : $(x + 1) + 1 = (x + 1) + 1$ and the jump is:

$$(x + 1) + y = (x + y) + 1 \quad \rightarrow$$

$$(x + 1) + (y + 1) = [(x + 1) + y] + 1 = [(x + y) + 1] + 1 = [x + (y + 1)] + 1.$$

Here we used three tricky equalities to achieve the consequence. The first used the addition rule, the second used the assumption, and the third again the addition rule just in reverse order.
So finally the exchangeability itself is $x + y = y + x$ which we already proved for $y = 1$ and the inheritance means:

$$x + y = y + x \quad \rightarrow \quad x + (y + 1) = (x + y) + 1 = (y + x) + 1 = (y + 1) + x.$$

The first equality used addition derivation, the second the assumption and the third the first member incrementing rule we derived above except in reverse order.
As we see it's a nightmare and it's quite doubtful that everything important about our operations could be proved by this. But that's it folks. All number theoretical theorems are proved as this.
As I said, it turned out that already the seven basic axioms are enough to achieve an underivability of the non theorems but this system obviously has undecidable statement because $x+y = y+x$ is not derivable without induction but this exchangeability is true for the real naturals and so the opposite of this claim shouldn't be derivable either if our seven axioms are true for them.

This means that the weakness of the Robinson system is mingled up with its strength.
How could we demonstrate its strength that it forces undecidable statements if it has trivial ones anyway? We have to show that adding new axioms to it the power remains even when the trivially undecidable statements disappear.
Now the "power" is due to the fact that it can simulate the derivation systems inside as properties.
To be more precise, for any derivable number set there is some Q (x) property that exactly those Q (n) cases are theorems for which n is in the derivable set. Unfortunately, adding new axioms some new cases might become derivable so the same properties will not necessarily represent the sets. We might argue that some new representation will emerge but that's basically then uses the new bigger system and so we don't have an automatic inheritance.
The solution is ingeniously simple. We have to regard those "rare" derivable sets that have derivable complement so are "dual derivable". What is obvious for these is that both the S set and its complement ¬S are representable by some Q (x) and R (x) properties. But what if it were also true that R (x) is actually ¬Q (x). This perfect representability would do the trick.
Indeed, then adding new axioms could not alter the fact that the dual derivable sets are represented, unless the two set of representing cases cross into each other which of course would mean that contradictory statements are derived and so the system would become inconsistent.
So we would need as assumption that our added axioms give a consistent system. But that's not a problem. This is really the minimal we can assume. Of course those derivable sets that do not have derivable complements can still be lost. Luckily, this doesn't matter either. The representability of these dual derivable sets is already enough to prove the underivability of the non theorems. In fact even the perfect representation by Q (x) , ¬Q (x) pairs is not important. It was merely a trick to get the inheritance.
So we get the incredible result that any consistent extension of the Robinson system has undecidable statement in it. This includes adding the infinity of the induction axioms if those are consistent. Which seems very much so since our real naturals obey them. This of course we can not prove inside, just see it from the outside.

So now I give the essential kernel of first Gödel's original argument and then of the newer and smarter one that we need after an inheritance of representability in the Robinson system:

Let T be a set of statements. We have the set of theorems in mind but we don't assume this.
T is trivially inconsistent if there is an A that both A and ¬A are in T .
T is complete if for any A exactly one of A or ¬A is in T.
T is incomplete if it is not complete and this can mean two things:
Either it is trivially inconsistent or not but there is some A that neither A nor ¬A are in T.
We call such A , ¬A pair as an undecidable pair in T.

Undecidability Argument:

If for every P ( ) property we can assign a |P| number so that there is a Q ( ) property that:

$$P(|P|) \in T \quad \leftrightarrow \quad Q(|P|) \in T$$

then if T is not trivially inconsistent then there is undecidable statement in T.

Indeed:
Using ¬Q as P

$$\neg Q(|\neg Q|) \in T \quad \leftrightarrow \quad Q(|\neg Q|) \in T$$

If both sides are true then T is trivially inconsistent which we assumed as not the case.
So both sides must be false and so ¬Q (|¬Q|) , Q (|¬Q |) are undecidable pair in T.

The word “derivable” that we normally use for single statements can be generalized for sets.
But just as we didn’t go into how the T statements were obtained above, we now also assume that the derivable S sets are simply given as a D collection.
This D contains the derivable number sets and also the derivable statement sets by any systems.
If both S and the complement $\neg S$ are in D then S is dual derivable or selectable.
The word selectable is very descriptive as dual recognizable too, which indeed means the same
The official nomenclature is different. It avoids the crucial extended usage of derivability for sets, yet quite confusingly, uses the word decidability for selectability.

Unselectability Argument:

If D is such that:

1.
For any S selectable number set we have a $Q(\ )$ property that:

$$n \in S \quad \leftrightarrow \quad Q(n) \in T$$

2.
For any S selectable statement set, the $S_0 = \{ |P| \, ; \, P(|P|) \in S \}$ number set is derivable (and thus of course selectable too because the $(\neg S)_0$ set is too and it is the same as $\neg S_0$)

then T can not be selectable.

Indeed:
Suppose it were! Then by 2. T_0 and $\neg T_0$ were too.
So using 1. for both T_0 and $\neg T_0$ there are Q , R properties that:

$$P(|P|) \in T \quad \leftrightarrow \quad |P| \in T_0 \quad \leftrightarrow \quad Q(|P|) \in T$$

$$P(|P|) \in \neg T \quad \leftrightarrow \quad |P| \in \neg T_0 \quad \leftrightarrow \quad R(|P|) \in T$$

The first line gives what we had in our previous proof and so it only proves that our assumption of a selctable T implies incompleteness but this doesn’t refute the assumption.
On the other hand the second line with $P = R$ gives:

$$R(|R|) \in \neg T \quad \leftrightarrow \quad R(|R|) \in T$$

which is impossible.

The mentioned incompleteness consequence for selectable T means in reverse that completeness implies unselectability. And this is possible. For example if we regard all true statements among the real naturals as T then it will be automatically complete but thus has to be unselectable.
So the true or false statements must be a non derivable set. In truth they are both!
The Unselectability Argument just gives that at least one of T or $\neg T$ sets must be non derivable. For usual T that comes from derivable axioms and logic, the derivability of T is trivial and so we get the result that $\neg T$, the non theorems are non derivable.
Then this with consistency implies incompleteness by the heuristic indirect argument as follows:
If completeness were true then the negation could be added to the normal derivations to get a derivation system for the non theorems.
This incompleteness of course with consistency means undecidability.

Finally we come to the result that could be called the crown jewel of the whole non derivable complement idea that I stressed so much. I simply said at the beginning that “most” derivational systems collect such sets. But this “most” was not only imprecise but strange too, considering that most classical systems have a complementing one and thus produce dual derivable or selectable sets. The evens and odds, the composites and the primes are all dual derivable. These are very primitive systems could I argue but then where do the complex ones start and how.
Gödel’s original approach was only showing that most axiom systems have to have undecidable statements. Many realized soon that the underivable complement is the basic heuristic idea behind this fact. But Turing’s machine approach was the clearest demonstration of this.
Unfortunately, the underivable complement was demonstrated through the universal machine idea or more precisely by the machine that shares input with every other. So actually we have a single example for what we claim to be typical, the complement being underivable.
This crown jewel widened the measly single example of this non derivable complement of the self referring universality into an arsenal of concrete examples. It brought down the veil why the classical systems were primitive and showed what we have to do to obtain underivable complement for sure.
Amazingly, the proof of this crown jewel is not hard and thus the moronic Formalist tendency regards it as a side result. The fact that neither Gödel nor Turing discovered it in the thirties and then even in the fifties Kleene missed it in his “bible” the “Introduction To Meta Mathematics”, shows that vision is the true evolution of mathematics not merely new theorems.
I compare this crown jewel of computability called “Rice’s Theorem” with the crown jewel of Set Theory, the “Well Ordering Theorem”. That is also fairly simple to prove but its vision reveals how the set collection method by explicit properties is able to avoid our timely collection vision.
The trick to get Rice’s Theorem is to realize that a seemingly quite reasonable assumption would imply that all derivable sets have derivable complement. This of course is refuted by the existence of Turing’s universal machine or more precisely by the machine that shares input with all others.
And so our “reasonable” assumption is actually false.
Then the falsity of this assumption will imply Rice’s Theorem at once.
So what is this seemingly reasonable claim?
To get it very visually, we should regard the big picture first.
The basic result of Turing’s program idea is that texts can collect texts.
This seems to be true in a very primitive level already. We say “the primes” and this collects the prime numbers which are actually indeed just texts using the ten digits as alphabet. So an initial problem is that the alphabet of the collector and the collected are different. The bigger problem is that this naming “the primes” is not an actual method. But we can replace such ad hoc naming with some formal system that derives the texts. The initial problem still remains that the alphabet used to express a formal system is different from the alphabet in which the systems collect texts.
Turing’s input idea solved this because a machine’s table can be written into as part of an input.
The price of this success is that the systems that collect texts are pretty over complicated.
They are not only machines that work deterministically as opposed to derivation systems, but then such machine must be used as a recognizer for all possible inputs using two tricks.
First to list all possible texts by length alphabetically and then list from these the recognized ones in dovetailing. So we try out all possible inputs increasingly for increasing many steps with returning to earlier ones and thus avoiding to be stuck at the failing ones. Of course, the actual order of the recognitions will not give the recognized texts in increasing lengths. In fact, this order seems to be irrelevant to the collected recognizable set. Yet now we’ll use this order with an even stronger emphasis. So we’ll use special lists as our effectivity forms.
For a Turing machine the order of operation could be very well visualized by displaying the states that it goes through. We would get a different list of states for every chosen input and these lists could be infinite if the halt state is not reached or finite if it is reached.
For our modified input collecting machine by the dovetailing the situation is different.
We wouldn’t have a dependence on possible inputs and the halt state could be reached again and again. To be more result oriented toward the collected texts we could print out these recognized inputs at every halt state instead of the boring halt message. On the other hand, at the other non

halt states we should be boring and just display a same special empty symbol ∅ to show that a non collected state has happened. So actually we simply regard infinite lists of texts that contain a special ∅ symbol and this also separates the normal t texts that appear as collected ones.
We can even allow as generalization that a t text could repeat which would normally not happen in dovetailing. This is logical because ∅ also repeats.
The allowed repetitions is really not much generalization because for any L list we can at once obtain an L' derivative of it that avoids the repetitions. Indeed, the already listed texts can be stored and compared with any new potential success and so we could omit the new.
The content of an L list is the B = [L] set of texts that appear in L between the ∅ signs.
The reason for using B is that later in Randomness we regard such texts as beginnings.
There the infinite texts are the main topic. Of course a finite text then should be called as segment but in randomness they boil down to the beginnings.
In the Randomness section we will also create for any L list an L* that is increasing in lengths of the appearing texts and we'll call this the dualization of L. The content of L* is not the same as of L and this is understandable already from the fact that if an L is increasing then it always has a complementing ¬L list. Indeed, we can list the missing ones up to any length.
Thus the usually non derivable complements mean usually non increasing lists.
Only dual derivable sets are contents of increasing lists.
This explains the naming of L* as dualization.
Though our lists are always infinite, their [L] contents can be finite if from a point we only have ∅ or we have new texts but they are all repeats of earlier ones. So this is the same as having a maximal length in our texts because up to this point there had to be a maximal.
So, the finite content lists should be called bounded while the infinite content ones as unbounded.
A drastic extreme case of the finite content is if there is no content at all just all ∅ signs.
To make a Turing machine that halts from no input is not hard and so this should be a possibility.
Now that we have a better formal vision of these infinite lists, we make the truly important conceptual shift and regard the texts as programs for such lists. For a fix framework of these infinite lists the list created by a t text as program is denoted as: [t > .
We'll call two lists variants of each other if they contain the same texts and similarly we'll call two t , s texts variants if [t > and [s > are variants, that is [[t >] = [[s >].
Or in short [t] = [s]. And this is the whole point. Text collection by texts that we can not achieve directly only through the listing with its seemingly arbitrary order.
Our original and main claim that most t texts collect a [t] set of texts that has non derivable complement means that the complement set of [t] denoted as ¬[t] is not a collectable set by text. So there is no s that ¬[t] = [s].
We'll get closer to this claim now but only through our list frameworks.
So now we can reveal the promised seemingly reasonable claim that would imply that all lists have complementing one:

If in a [> framework there were an E list that contains all t texts that [t > is empty but
E does not contain an s text and neither any s' that is a variant of s ,
then for any L list there were a ¬L list containing exactly those texts that are missing from L.

The wild goose chase to look for a t in our L list if it doesn't appear is an actual empty list.
This has a p program and the targeted t text can even become a parameter of p too.
So the E listing of all empty programs is a logical start. But unfortunately this means to allow all possible t and so then the successful searches where t does appear in L would melt in.
The crucial idea is to use the avoidance of the totally unrelated s' variants by E to avoid the contained t texts of L. And the trick to do this is to create for two L and K lists a list of all possible pairs of members taken from L and K. This should be denoted as L x K.
The pairs can be imagined as t + t'. But here we mean all members including ∅ and so these combined members can be ∅ + t' or t + ∅ or ∅ + ∅ too.
To make such L x K is actually the simplest form of dovetailing or the trick used by Cantor to list all fractions. We simply go further and further in the lists to make new pairs with olds.

For a fix such list multiplication then we can introduce a division by a fix t text.
Then this (L x K) : t will lists all those second t' members of the t + t' pairs that have the fix t as first.
For a t from L we obtain back the members of K while for a t not in L we'll find nothing.
So (L x K) : t is a variant of K or empty according to whether t is picked from L or outside.
If our list framework is good then this listing could be also obtained as [t* > with L and K regarded fix and only t as variable. So t* gives an effective translation of the t text into a t* program to create the (L x K) : t lists in our framework for every possible t.
Now we can choose K as the claimed [s > and then t* is for (L x [s >) : t and so:
[t* > is an s variant if t is from L and is empty if t is from outside of L.
Finally, we'll use the claimed E to filter out the s variants and only keep the t* empty list programs. Thus for t we get exactly the complement of L.
In details, we first regard a fix U list of all texts and then for all the n naturals we do the followings:
We regard the n-th member in E which is e_n and also the n-th member in U which is t_n .
We translate this into t_n*. We also store these e_n and t_n* texts in two memories and so we can see if any new e_n or any new t_n* already appeared in the other memory.
So in fact we could list $E \cap U^*$ the cross section or common members of E and the translated U list. But we don't list these rather their non translated t_n texts.
They are exactly the texts missing from L.

Now we regard a drastic strengthening of our impossible E list. Namely:
We assume an E* so that it again contains all empty programs and avoids an s but now it should avoid not only the s' variants of s but in general all variants of any r that is avoided.
First of all observe that this also means that for any r that is in E* we have all r' variants in E* too. Indeed, for any variant complete T text set, the complementing $\neg T$ is variant complete automatically. So this stronger E* should be called a variant and empty complete set that is not the set of all programs. And of course such can not exist either.

This implies at once Rice's Theorem:

If two T , $\neg T$ complementing text sets both contain some texts and they are variant complete, then both can not be derivable by machines.

Indeed, one of the sets would contain the empty programs and so it would become the refuted E*.

So if we can create T , $\neg T$ text set pairs that are variant complete, non empty and we know that one of them, say T is derivable, then we could be sure that $\neg T$ is not is not derivable.
And to create such T sets is quite easy. All we have to do is let any t text run as program to create [t > and regard a property that can be established after some beginning of [t > already.
This is automatically variant complete because we don't check t as program only its list content.
The only nuance is to check that there will be some texts collected and also some avoided.
Such beginning list behaviors are easy to tell. Lets see a few examples for number lists:
Whether a program will list the number 10 is clearly such. To list at least one prime or at least twenty primes is again.
But whether the number 10 will not be listed or no prime will be listed or no more than twenty is listed are not recognizable intuitively because these can not be assured from a beginning.
And indeed, these are the opposite sets of the previous ones and so their program sets are definitely not recognizable by Rice's Theorem.
Whether all primes are listed is again an intuitively not recognizable property because in a beginning we can see only finite many. But now the opposite, that is not listing all primes is also non recognizable intuitively because they all can still pop up after any beginning.
An even simpler case of this is whether a program will list exactly twenty primes.

Rice's Theorem can not be used for these intuitively dually non derivable program sets.
The most obvious case when Rice's Theorem does work is of course the empty programs.
The t texts that [t > are empty can not be listed because it would be an E and also an E*.
But this also follows from Rice Theorem because the t texts where [t > are non empty are recognizable.
Collecting all programs that list the number 10 will obviously contain all those programs that list only the number 10 and this suggests a paradox. We could collect all those programs that contain a ∅ and then we would collect all those programs too that contain only ∅ and so we would collect a variant complete set containing all empty programs. Seemingly this is a forbidden E.
What we forgot about is the avoidance of some s' programs or in the Rice's Theorem version that both T and ¬T must contain programs. Containing a ∅ is true for all programs.
A bit weaker condition than being empty is to have a finite content.
Remember that the opposite, that is having infinite content is actually having arbitrary long texts, what we called being unbounded.
It's easy to show that the programs for only these unbounded lists can not be listed.
A smart way to do this is to prove that for any list of such unbounded lists there has to be some missing from our list. So we regard simply a list of lists. We must allow repetitions just as we did in the generalized text lists because different programs can produce same lists.
So let L_1 , L_2 , . . . be a list of unbounded lists.
The k-th text in L_n could be denoted as $t_{n,k}$.
So the first text in L_1 is $t_{1,1}$.
Increase $t_{1,1}$ with a symbol to get a $t_{1,1}*$
Now go in L_2 till the first t_{2,k_1} text comes that is longer than $t_{1,1}*$.
Increase this again to get $t_{2,k_1}*$.
Find the first text in L_3 longer than $t_{2,k_1}*$ and it is t_{3,k_2} .
And so on, we get an obviously increasing L list that can not be identical with any L_n .
The first text of L is already longer than the first in L_1 and so all texts in L will be longer and so that first text is missing from L definitely. But then the found longer text of L_2 will be missing again because we skip that length completely. And so on, all L_n lists will have one definite member that can not be in L.

Now we prove that even just a partial listing of some unbounded lists can be impossible.
Namely, if that list would contain all unbounded lists that have complementing lists.

Here we will actually demonstrate two missing such lists at once.
We copy the first two different long texts from L_1 onto two A , B lines as first members.
Then we go in L_2 till we find two different long texts both longer than any of the two previous ones. And we place again one on A and the other on B.
And so on, we always pick new longer texts and place them separately on A and B.
The A and B lists are again both increasing and now this is important because it means that they are lists having complement lists. Being increasing also means that they are unbounded but their complements may not be such.
Also observe that A and B can not have common member.
But the L_n lists will contain common member with both A and B.
Thus neither A nor B can be identical with any of our listed lists.
So A and B are unbounded lists having complementing lists yet not among our listed lists.
To be perfectly precise we should have told how we separate the selected two texts into A or B.
We have to make this effective, say A containing always the shorter and B the longer text.

A trivial consequence of this result is that we can not list all those unbounded lists that have complements.
The fundamental construction of Turing, the machine that shared input with all machines and thus had no complementing one, also the Undecidability and Unselectability Arguments had the same common “diagonal” idea that Cantor used first and we used above too. So we might think that in more general, a listing of all lists that have complementing lists should be impossible too.
But in all these arguments the complementing pairs were correlated in some sense.
And indeed, it’s easy to see that the ordered pairs of all complementing lists is not listable.
But quite amazingly, the unordered pairs, or equivalently the lists alone that merely have complementing pairs can be listed. Such list is very hard to demonstrate and it obviously hides the complementing pairs so much that we can not obtain a diagonal contradiction.
Much easier is to show that the above proved unlistability of the unbounded lists does not follow from Rice’s Theorem because the bounded that is finite content lists are also unlistable.
In fact, there is a generalization of Rice’s Theorem to characterize the listable lists and it exactly relates to the bounded lists.

And now finally some critical words.
As some kind of “counter attack” against the heuristic “program” vision of Rice’s Theorem, a new concept became the so called “numberings”. This continues that already happened in Kleene’s approach. Number functions as replacement of Turing’s framework. These were much more precise than talking about programs but the true arguments were program visioned.
Now the programs are actually replaced by numbers and regarded not as the collectors of the sets rather as merely attached labels to functions and sets. Of course a lot of special requirements can be then made about these so that they behave like programs. And thus abstract result can be obtained without referring to programs.
I don’t claim that this is intentionally created to deceive the outsiders and I am sure that some mathematicians can think in this abstract field without even going back to programs directly.
It is still a lie. And that is why this approach will not make the real breakthroughs. For that we have to go not above with abstractions rather even deeper to the bottom.
A less substantial deception is the following:
Soare championed a long crusade to rename all the old “recursive” adjectives as “computable”. This in itself is a half baked goal because the main fault of the old approach was the lack of the derivability vision for sets and rather starting with selectable ones only. So the basic concept of derivable or recognizable or listable was overcomplicated as “recursively enumerable”.
This craziness still remains as “computably enumerable”. In fact, the word computation hides the main important feature of the infinite possible runnings. If a computation goes on for ever then it must use infinite searches. So we use a generation.
But the real point is deeper than this. Namely, that the effective collection of sets, is not only the basic intuitive vision which I call derivability, but also the main results are about complementing contents. So the particular list orders are just our bad approaches because we have no systems that would grasp the contents directly. In short, a good framework of derivation systems does not exist.
Gödel chased such for years but never said why he was unhappy with the unnaturally over mechanized concepts of fix ordered derivation systems.
Turing of course boldly went in the directly opposite direction because he regarded the machines being deterministic by nature. It is a falsification of history to claim that Gödel at once accepted Turing’s method as the final solution. No! He simply reluctantly accepted that a preferable formalization of effective collections did not succeed.
But now we live in a new brain washing age with stupid bureaucrats controlling information.
Neither Gödel’s nor Turing’s groundbreaking articles would be published in any journal today.
And of course Newton’s Pricipia would be rejected too.
These are writings that not simply proved already raised questions rather provoked new visions.
They would be labeled today as incoherent ramblings.

Part Three : Randomness

1. The Fundamentals

The original basic vision of Set Theory was the collection principle. For any P(x) property we should have a set denoted as { x : P(x) } that collects exactly those x sets that posses the P property. As it turned out, this principle is contradictory. So the collections had to be pruned and the allowable collections as set existences became the main set of axioms.
But at the same time a second, earlier unrecognized though widely used principle also surfaced. Most mathematical arguments start like this: Let A , B , C be this and this kind of objects. Like three corners of a triangle. Then we examine these and derive some results.
The "kind"-s that we assumed about them, are described by some properties and we know that they exist. So this picking of finite many and well denoted objects is a method of logic that in precise form starts with the assumption of these kinds as existences and then introducing new variables as representatives for each. These are actually temporary names for the length of our arguments. In fact, this method was the guiding idea of Hilbert's axioms for logic.
Now a trivial generalization of this method would be that if we have an R (x , y) relation and we know that for every x value we have some y that R is true, then we should be able to pick a y for every possible x. Unfortunately, we can not say this strictly in logic because this "for every x having a y" is not a logical basic formation. In truth it should be, but logic was not extended in this direction because Set Theory was ready for this job. Namely, offering functions.
For example, we say: pick for every n natural a bigger N. Here R (x , y) is x < y.
This picking is a purely logical step but what we claim is an f function so that N = f (n) > n.
The form of f as such is not enough because as I mentioned above, some formations can be contradictory and so we needed existences as claims when we have actual sets. So here too, we need an axiom that says that such f exists. In general, we can even avoid the use of an R relation because all objects are sets anyway. So the claim can be formulated as this:
If an S set has only such s elements that are all non empty, that is do have some elements, then there is a c function that picks as value for every s element of S an element of s.
In symbol c (s) ∈ s. The previous f (n) > n for example can be then obtained as follows:
We regard for every n natural, the numbers above, that is { x ; x > n } as the s (n) sets.
So a c exists that picks from every such s (n) a c (s (n)) and it is then f (n) = N.
Of course, this new Axiom Of Choice is only needed in arguments where we can not give some other way of showing that such f exists. The most obvious example of when we do not need this axiom is if S is such that it has a common element as choice in all s elements.
Indeed, if e ∈ s for every s ∈ S then a possible choice function is f (s) = e for all s.
In our example of s (n) = { x ; x > n } this doesn't stand because the sets of numbers above bigger and bigger numbers don't have a common element and yet the Axiom Of Choice is not needed because we do have an other obvious f , namely as f (n) = n + 1 for example.
Observe, that the common e choice also means that maximum only one s can be that has only this e choice, because sets can not have repeated elements and so the singular {e} set can only be one element of S. All the others must have other possible choices beside e.
As it turned out, such at least dual choices implies that the set of all possible choice functions can not be equivalent that is one to one assigned to the set from which we made the choice functions. This then is interpreted as this set of all choices being bigger in cardinality.
This is the real start of Set Theory and it didn't need the Axiom Of Choice.
And yet, later the Axiom Of Choice sneaks back in more subtle ways.
The dual choices brought not only this amazing result for the choice cardinality but can also be continued to distinguish more knowable choices. Indeed, the all e choices was enough to see that we have possible choice function but we can obviously create other known variations too.
To make more such, we have to use those elements of the s elements that distinguish them.
These give a structure of the S set. The simplest infinite structure is the sequence of the natural numbers and so we get as simplest structure with dual choices the following universe:
{ { 1 , e , o } , { 2 , e , o } , . . . }.

Here e is the mentioned basic common choice, o is an other, to have a real choice and the numbers are the structuring of the elements.
To simplify the notation, we can use 1 for e , 0 for o and we can avoid the naturals if we place the elements not as a set rather as a sequence where the place is the hidden natural:
{ 1 , 0 } , { 1 , 0 } , . . . A choice function then becomes an infinite sequence of 1 or 0 -s.
The earlier fix choice function that avoided the Axiom Of Choice is thus 111111 . . .
But we can create other sequences as choice functions like 00000 . . . or 1010101 . . .
The field to find all such concrete choice functions is the field of Effectivity and it had many frameworks. Every framework defines so called "systems". A system is described finitely but can collect infinite many objects. The objects can be natural numbers, or consecutive finite blocks of choices. In a classical mathematical approach the number object was more natural but for randomness the finite choice blocks are better. It is then a perfect coincidence that using machines as systems also prefers objects as choice blocks. These can be finite texts on some alphabet but could also be finite more dimensional data sets like digital pictures made from fix colored squares. The alphabet or the possible colors are the possible choices, the choice set. Dual or binary choice set means texts made from merely 1 and 0 or pictures made from black and white little squares. That this indeed can produce pictures from a distance is obvious if you take a magnifying glass and look at a newspaper.
The most logical idea to define effectivity should be to generalize mathematical derivability.
The derivable theorems then become merely derivable texts without the presumptions of truth and falsity. This is perfectly analogous to games where the texts are situations and the derivations are the game histories. I traveled this line in the second part of this book to reveal why the non derivabilities are actually very logical. The crucial point was that instead of the individual derivability of singular objects there is a derivability of a whole set of objects. Namely these are the collection of all the objects that are derivable by a fix system. And I prefer this word usage, that is to call these effectively collectable set of objects as derivable even though this framework was not the successful in revealing the real situations.
The best of the more revealing effectivity frameworks became Turing's idea that uses singular choice alterations only. This corresponds to how an elementary school child is taught to add, subtract multiply or divide large numbers digit by digit. And this is also the only way how a machine could perform these. No understanding just mechanical precision. It's amazing that this blind and least mathematical framework became the most suitable to reveal the new main idea that the seemingly so logical derivation systems keep hiding, namely the concept of programs.
So generalizing the method of the digital calculations of the four basic operations to any possible new systems, we get the followings:
We rewrite any single square and then move to a neighboring one. At a text this means that we rewrite any symbol then move left or right while at pictures we recolor a square and then move to one of the four neighboring ones.
The system that tells how to do these local alterations is a simple finite set of transition steps called the "transition table". This tells that after every possible read symbol and so called machine state what new symbol should be written, which neighboring square we should move and what new state the machine enters. So the introduced machine states are the essence behind.
The structures most obviously suited for these alteration systems are infinite in all directions.
For texts it means a two directionally infinite line of memory squares, while for pictures a full plane of memory squares. A choice function is a fully loaded memory line or memory plane.
The alteration could start from such and from any chosen square. This is an amazing idea but could only be called relative effectivity because the initial line or plane was not effective.
To define effectivity, we should start from a trivially effective initial data set. Simplest is to have all memory squares with the same choice which then could also be called "empty". A more flexible generalization is to regard empty initial set but to allow a finite block too as input.
It can be shown that regarding only one directionally infinite memory line, that is using the naturals as structure for the choices can obtain the same effectivities as the full line or plane.
Also, a binary alphabet like 1 and 0 or black and white colors is sufficient and even as empty initial set we can use one of these like all 0-s or all whites.

The transition tables as machines are the simplest computers that can do all that real computers can. Turing was very conscious of this in a time when there were no real computers yet.
In fact, the Turing machines could be called as a “crawling computers”. This expression refers not merely to the fact that local alterations must crawl square by square but has a more surprising meaning how such crawlings can be successful as programming. Indeed, the transition table is just a very minor part of how we can alter the memory line. A table containing only hundred transitions could use a million long data segment of the line as program. This then requires that this memory segment is kept safe and we crawl back and forth to this to alter the rest of the line. Real computers alter the data parallel to be fast.
In a Turing machine even simple programs would take millions of years to execute. But here in this theoretical framework time is irrelevant. On the other hand, in this seemingly inferior framework we have the infinity of the line so an infinite memory that no real computer has.
So actually a Turing Machine can do much more than any real computer can.
Of course, all other effectivity frameworks target this wider content that uses infinite memory.
The Effectivity Thesis is the claim that all effectivity frameworks are equivalent. Meaning that whatever alteration of an object we can achieve with any system of one framework, we can achieve it with some systems of the others too. We might think that here the “achieve” means any particular alteration of an object like text or picture. But this is not enough!
Indeed, a system’s capacity is how it alters any initial text or picture.
And so to “achieve the same” should mean that for any initial input we can get the same alteration. This makes the concept of the “collected objects” by a machine relevant too.
These could be defined either as the altered objects achieved from all possible inputs at a certain halt state, or ignore these results and rather collect those inputs that lead to halt.
The Effectivity Thesis is not provable because we do not have a trivially universal framework.
The frequently hidden application of the thesis is by using some intuitively effective steps and assume their validity in some concrete frameworks. This can always be replaced by checking how our concrete used frameworks are equivalent.
Surprisingly, right at this start and without going further into effectivity, we feel that there is an other extreme for choice functions, opposite to the effective ones. Namely those that are done randomly. But most amazingly, withdrawing our artificial creations and so having only choices randomly, is not a negative of effectivity. Of course, effective choice functions are automatically non random but not to be random, the choices don’t have to be effective. As simplest example, it is enough if every second choice is effective say a fix 1 and the other ones are random.
So a choice function looses its randomness already if any effectivity can be detected in it.
Obviously, effectivity must enter somewhere into the definition of randomness but the start should be to see how this much wider “detectability” could be defined.
A “strangeness” is any property that shouldn’t be true for any random choice function. These are the ones that hide some effectivity. The negative of a strangeness is an “expectability” which is therefore true for all random choice functions. The amazing thing is that we all have dozens of intuitive strangenesses and expectabilities. But the real question is the reversal!
Can we collect some strangenesses that together form a universal strangeness? That is, any non random sequence would obey one of these. Or can we collect some expectabilities that together form a universal expectability? That is, any sequence obeying all these would automatically be random.
An even bigger surprise than our a priori intuitions comes out only by logical examinations. Namely, how such intuitive strangenesses or expectabilities can imply each other.
Then a third newest surprise became that even though we can get same sequences defined as random by very different intuitive ways, we can also define a whole arsenal of different variations. So as usual, a hidden new world got revealed too.
To decide whether there is an absolute concept of randomness is therefore quite difficult yet.
The real beauty would be if at the end all the original players, Logic, Set Theory, the Axiom Of Choice, Effectivity would interrelate through Randomness. This is still only the future!
Now a more precise line of the earlier concepts:

A choice universe is any S set that contains in every s element a fix choice set also called as the alphabet. The structuring of S is done by the other differing elements inside the s elements but we envision this outside. We also take the choices outside and so we envision repeated empty memory cells in the elements of a structure like on a line or in a plane.
Then a choice function is actually a filing of this infinite memory with arbitrary choices of the alphabet. So a choice function could be called a concrete universe or just universe for short.
A random choice function is a random universe. Like an infinite random text on a line or infinite random coloring of a plane.
A window is any finite many points in our structure.
A window set is any collection of windows. The simplest collection could be by the number of elements. An other simple restriction on the line is that the window points are consecutive which is also called as a segment. Usually this is what we would regard as a window subjectively but we use the word window more generally. So every second or third points used is also a window. In the plane a window set could contain all ten by ten squares.
A window choice is making finite many choices.
A property is any collection of window choices. With $\{1 , 0\}$ choice set, to have more 1-s than 0-s is a property that collects all window choices where this fact is true. No assumed window set was used here but most of the time we do include window sets as restrictions in our properties. For example, on a line we can collect only all segments that have more 1-s and so we get a smaller property. In the plane we can collect all squares that depict Mickey Mouse from a distance with using black and white tiles.
If we have a structure with coordination like a half line or a full line or plane with origin, then we can use in our property the feature of starting from the origin. At a sequence these are the beginnings. This going from an origin can falsely project a sense of time. But our true final plausibilities about the random choices have no time involvements.
Infinite many simultaneous choices are exactly the same as done one by one from a start if we have an origin at all. And yet, as we proceed we will use such false timely plausibilities.
The simplest such is the Law Of Restart and it claims that in a random sequence from any point the continuing sequence is again a random one.
The true spatial vision is simply that in a random universe, altering any window choice the universe still remains random.
The second falsely time involving plausibility is the Realistic Murphy’s Law.
Unlike the cynical one that claims that a buttered bread always falls on its buttered side, this merely says that if a feature has a positive chance and we repeat trials of the feature then it will succeed sooner or later. Falling on the buttered side has half chance so if we keep on dropping our bread then sooner or later it will land on the buttered side. Similarly a coin must sooner or later land on head and also on tail.
Amazingly, applying the first law that randomness restarts, we can conclude that any feature with a positive chance must repeat infinitely. Indeed, after it succeeds, the continuing trials are again a random sequence and so the Realistic Murphy’s Law applies again and again.

The flips of a coin are independent and so the chance of a prescribed segment happening can be obtained by simple multiplication. For example, obtaining three heads in a row has $\frac{1}{2}x\frac{1}{2}x\frac{1}{2} = \frac{1}{8}$ chance and this is understandable also by the general rule that a feature’s chance is always the ratio of the succeeding cases over the number of all possible cases.
Here we have one success hhh and eight possible cases: hhh, hht, hth, htt, thh , tht, tth, ttt.
Any other prescribed triple would have this same chance so three heads is not more special than for example htt.
A ten long segment has only $\frac{1}{2^{10}} = \frac{1}{1024}$ chance but a trial sequence of ten flips can be done in less than ten seconds and so 1024 trials for such ten long segments can be accomplished in about three hours and so in that time we can expect any ten long combinations to occur.

As a contrast, if we go up to thirty long segments than our whole lifetime is insufficient to expect a success of any prescribed one.
Of course, these "chance expectabilities" from the number of all combinatorical possibilities men nothing sure. Indeed, a head or tail means two possibility but two trials does not guarantee a head and a tail. We see many ten long trials without a head or without a tail. Of course, twenty or thirty consecutive trials without a change would be much stranger but this is merely because we are well equipped with an "a priory", that is born estimations of these chances in human times.
This is amazing philosophically. But even more important is that, we do have a parallel logical "a priory" set of rules which tells us that in spite of these real time expectabilities the Realistic Murphy's Law is true too and so the minutest chances that seem very unlikely to occur must come about if we wait long enough. So everybody knows that even though winning a lottery has a minute chance, if we could play week after week for trillions of years then we should hit the jackpot. And we also know that then again infinity is there so we should win again and again.
This above mentioned combinatorical expectability had its probability inducing paradoxes and this is how probability calculations started. Throwing two dice has 36 cases and so we can only "expect" a double six after 36 throws which of course is not a "must" but how these chance expectabilities should be combined was that started probabilities. I don't go into this line now just as those early probabilitists didn't go into the question that I peruse now, namely how the particular infinite trial sequences should behave. For us now the mentioned plausibility of the Realistic Murphy's Law means that if we imagine consecutive fix n long windows in a random sequence then each can be regarded as trials for a chosen outcome combination of the segment.
This has $\frac{1}{2^n}$ chance which is positive and so repeating it infinitely that is checking these n long windows, we sooner or later encounter success of the chosen combination.
Even more amazingly we then have to encounter the outcome combination infinitely too.
To make this situation even weirder we can regard not a dual outcome trial, rather one with much more possible features. A dice has only six sides with each having $\frac{1}{6}$ chance but to hit a chosen letter on a keyboard has about $\frac{1}{100}$ chance if we do it blindly.
If we tie a monkey to a chair and train him to hit the keyboard, it would also create an infinite random sequence. If we again regard a long window say a billion many keystrokes then we have truly a miniscule chance for a prescribed billion letter long text to come about randomly and so also by the monkey. Nevertheless, by the Realistic Murphy's Law these also have to come about sooner or later and in fact infinitely. And so, actually the full Bible will be typed by our monkey sooner or later and then infinitely too. This "Borel's Monkey" thought experiment was made at the beginning of the last century with typewriter of course. By the way, we claimed the infinite occurrence of the Bible in our fix windows but the "first Bible" most probably should have occurred even before, where it was not starting in an exact window rather in between and overlapped into the next.
In a much better spatial vision these paradoxes are tamed a bit. Indeed, they simply mean that if in a universe in any sequence of non overlapping windows the chance of a property is above a positive value, then in a random universe we will see infinite many of the windows where the property is true. And so, for example flying over a random infinite chessboard we will see arbitrary chosen pictures like a Mickey Mouse to appear infinitely. A hidden proportionality was assumed here that implies the chances to be the same. Using same long windows we don't need this and then any fix outcome as case is a fix chanced property anyway. This is also the reason for the claim that in alternate universes we all exist already. Indeed, we are made from atoms with finite choice combinations. So here the spatial vision becomes paradoxical again too.

2. The Law Of Stopping

The other two famous expectabilities involve not segments like the Realistic Murphy's Law rather beginnings. These are the Law Of Large Numbers and the Law Of Dual Oscillations.
Before even telling them we have to realize that to claim some tendencies about the beginnings is actually mystical. Indeed, for the one by one trials to bring about a tendency in the beginnings means that these outcomes somehow know or are influenced by the past.
This already dictates that we should eliminate these and replace them by something more rational. And this is perfectly possible by laws that avoid the coordination.
Then it turns out that it is not the widening of the beginnings rather the mere lengthening of the windows that causes our claims. We have also two such universal expectabilities:
The Law Of Stopping claims that very low chanced properties can only become true finite many times in our windows. The Law Of Occurrence claims that some very high chanced properties must come about in one window. The condition of this law, that is the high chance might seem simple as 1 total but this only gives certainty for excluding choices.
In contrast, for the Law Of Stopping we need no restriction other than the low chances.
On the other hand, here having 0 as total is obviously a crazy opposite of 1 and instead some kind of smallness is required. Also, we might think that the Law Of Occurrence is similar to the Realistic Murphy's Law and thus will lead to an infinite occurrence.
But that used new windows that imply independence while here we can have overlapping windows and instead we restrict the outcomes to be excluding. This means not allowing continuing ones and so only single occurrence can be claimed.
These problems can still not explain why the two fundamental laws surfaced so late!
The Law Of Stopping by Solovay and the Law Of Occurrence by Kurtz.

The chances in the windows are $\frac{k}{m^n}$ where k denotes the number of cases with our property in a w window, m denotes the alphabet size and n denotes the window size.
Indeed, we have m^n many possible n membered choice sets from m many choices.
By the Law Of Stopping a property will be stopping because our chance function is not like the chance of a picture that is proportional and thus remains same for increasing sizes. Instead, it becomes less and less drastically. Both the "less and less" and the "drastically" can be defined precisely on its own and surprisingly it is independent of the structuring of the universe.
The chance values in different windows can repeat but as I mentioned, elements can not repeat in a set. So instead of the chance value set, we regard a p value function on a W window set.
Our value conditions will be quite general for any p positive real function on any W set:
p is concentrated if fore every positive ε value there is a W_ε finite subset of W so that outside W_ε that is in $W - W_\varepsilon$ every p value is under ε.
p is well concentrated if in $W - W_\varepsilon$ not merely the values are under ε but their total is too.
The Law Of Stopping then claims that:
If a property is such that in a W window set the p chance values are well concentrated,
then in a random universe only finite many windows in W have the property.

Observe that the windows are themselves finite sets and the well concentration involves finite sets that are however finite sets of windows. To see even clearer what we claim:
Let the outside windows in $W - W_\varepsilon$ be $w_1, w_2, \ldots$ with p values $p_1, p_2, \ldots$
We claim that if $p_1 + p_2 + \ldots < \varepsilon$ for all ε values then we can only have finite many windows in W where the property is true.
We already used the simple multiplication law for independent chances. That calculated the chance of an "and". A corresponding addition law applies for excluding chances to get the "or".

The six sides of a dice are such and indeed the chances of say having an even throw is the sum of the three $\frac{1}{6}$ values giving $\frac{1}{2}$ that we expected.
Consecutive trials are of course non excluding and so we may wonder what relevance this law could have. Actually this law is much better because non exclusive possibilities can only reduce the chances of an "or" and so we have a law that is true for all trials:

$$p(1 \text{ or } 2 \text{ or } \ldots) \leq p(1) + p(2) + \ldots$$

There is an other universal inequality law, namely that if a claim implies an other and this consequence has a p chance than the claim that implies can not have more chance than p.
Now the opposite of our claim that we would have infinite many successes implies that outside any finite set of windows at least one success happens. This "at least" means "or".
So the chance of infinite success is not more than the chance of the "or" outside any finite set of windows. But this "or" is not more than the sum of the chances that is $p_1 + p_2 + \ldots$
But these become arbitrary small and so the chance of the infinite occurrence is not more than any arbitrary small value and thus it can only be zero.
Now comes the final step to say that therefore this infinite occurrence is impossible.
This is the most unjustified actually because we know many outcomes that have zero chance and yet can happen. For example to hit an exact point on a dart board.
Why our window trials are different is actually still a mystery.
Probability Theory doesn't make this assumption and so it does not claim that these properties must stop. It merely claims that those sequences where they stop is a much bigger set in measure than those that don't stop. Or equivalently that they stop with probability 1 .
We achieved that in a random universe any window property that has well concentrated chance values must stop. The more plausible yet falsely timely version is a reverse of the stopping from the property to the randomness. So it says that that a random sequence must stop in the property if the chances are well concentrated. In fact, even the concentrations are falsely time envisioned as diminishing or fast diminishing. Indeed, thinking in a $p_1, p_2, \ldots$ sequence of values from the start, for every ε value we have an N so that after this as index, that is for $n > N$, either the p_n values or their total must be under ε.
Strangely, there is an alternative of the well concentration or fast diminishing in this timely vision that destroys the timely vision itself and reveals the true spatial nature of our claims.
Namely, to have these external or at sequences "end" sums become arbitrary small is equivalent to simply that the full total of the p chance values is finite.
Indeed, the partial sums up to more and more members approach this finite value and so the leftovers have to diminish. Also in reverse, if the sum is not finite rather infinite then the leftovers are always infinite too. And a total sum is irrelevant of the order we add them up.
Surprisingly, this intuitive final fact is only true because the values were positive.
With plus and minus members an infinite sum is ambiguous. Just think of this:
$1 + (-1) + 1 + (-1) + 1 + \ldots \quad 1 - 1 + 1 - 1 + 1 - \ldots = 0$
But we can rearrange the members into $1 + 1 - 1 + 1 + 1 - 1 + 1 + 1 - \ldots =$ infinite.
Finally, we can ask why the Law Of Stopping should be a suitable randomness criteria.
We do feel that strange beginnings should stop in a random sequence and we do have instant plausibilities about concrete beginning properties being strange too! But sadly, we don't see that these always mean fast diminishing chances and that actually this is the cause of stopping.
This is exemplified by the following paradox: Lets regard the naturals up to ten million.
Now lets exclude a certain digit say 7. What percentage of numbers remain?
We might think that the result should be about ninety percent because we only excluded one digit from the possible ten. The shocking truth is that less then half of the numbers remain!
Up to ten we indeed have 90 percent remaining because we exclude only 7.
Up to 100 we exclude all the seventies that is already 10 percent and from the remainings we exclude all 7 ending ones that is an other 10 percent of these. So actually the remaings are 90 percent of ninety percent, that is $.9 \times .9 = .81$ portion. Continuing similarly we get that up to ten million the remaining portion is $.9 \times .9 \times .9 \times .9 \times .9 \times .9 \times .9 \cong .47$.

Not to have a single 7 for ever in an infinite decimal is obviously a strangeness and these are less than an arbitrarily small portion of all decimals. They are what we'll call a nil set.
But our example was confusing because the strangeness was hidden under the base ten system that is having a ten member choice set, plus the beginning property of not having any 7 was automatically true for all sub beginnings.
Regarding more flexible beginning properties of decimals, the essence remains the same:
If a property is strange to occur infinitely, then its chance among the longer and longer finite decimals must diminish fast. Or in a spatial form, the chances must be well concentrated. Or to be even more precise, the total chance of all the beginnings with the property must be finite.
We could regard for such beginning property an infinity of new beginning properties that require merely that the previous property appears k many times among the sub beginnings.
Infinite occurrence of the original property obviously means that these new properties occur at least once but with higher and higher k values. We'll express this by saying that: "these new properties cover the sequences". And now these must only diminish that is be concentrated. This will be proven precisely and in reverse too. Thus the Law Of Stopping has an alternative meaning. Instead of claiming that a random sequence stops having beginnings with a fast diminishing property, it means that a random sequence stops being covered by a sequence of diminishing properties. In fact, the Law Of Occurrence is also equivalent to a special case of stopping being covered. Namely, by a sequence of diminishing properties each with only finite many members. So the concept of random sequences stopping to have beginnings that obey something strange, is the universal vision of all expectabilities.

3. The Law Of Large Numbers

This became a household expression in the 20th century but like at many other fashionable expressions most people have not a clue what it means. So lets start with that:
If a feature has p chance then in our trials up to infinity, regarding in the longer and longer N long beginning the n number of successful outcomes, the ratio $\frac{n}{N}$ must approach p.
A real meaning for the Law Of Large Numbers could be to define the probabilities.
Instead, we usually get the p values from the physical features. Like a coin having half chance for the head and tail and a dice having a sixth chance for all numbers.
But then the purely mathematical approach of Probability Theory repeats this same reversal and starts with assumed probabilities rather than the Law Of Large Numbers.
Even more strangely, then these probability books do get to laws by this same name but do not say what we claimed. Instead some much more complicated limit theorems. So what's going on?
These limit theorems regard the sequences as new events and do not avoid the strange ones. Therefore they can only assume that the strange ones are rare or have 0 chance while the random ones have 1 chance. Without of course defining these, they rather replace the validity of the approach of the success ratios to the assumed p values merely by 1 chanced claims.
This left many mathematicians still in the search to define the random sequences and that's why this whole subject of randomness survived in spite of the success of Probability Theory.
The leader of this originally lonely pack was Von Mises. But his unfortunate obsession was exactly the Law Of Large Numbers. In fact, this ruled the minds for fifty years.
And yet I can show in a minute why it is so misleading.
It seems to claim a tendency about the beginnings, an expectability of them like the reoccurring outcomes were in the Realistic Murphy's Law. But those were logical by the open future and the new restarts. Here it would mean that the outcomes are influenced by the past or know the past.
But this is an illusion! In truth, this law claims the stopping of certain beginnings. We just have to look closer what the claimed approach of $\frac{n}{N}$ to p means.
First of all, it is not merely a getting close claim rather a staying close, in other words:

For every ε positive value we must have an M number so that if the N long beginning is longer than this M, then our $\frac{n}{N}$ success ratio is closer to p than ε. Or in even shorter form:

For every ε there is an M so that for every N if $N > M$ then $| \frac{n}{N} - p | < \varepsilon$.

The absolute value simply meant the positive difference between the ratio and p.
The cat will start to come out of the bag if we ask what really the implication at the end means.

It means that it is not true that $N > M$ and yet $| \frac{n}{N} - p | < \varepsilon$ is not true.

This last not true of course means $| \frac{n}{N} - p | \geq \varepsilon$ and so our claim is:

For every ε there is an M so that for every N it's not true that $N > M$ and $|\frac{n}{N} - p| \geq \varepsilon$.

We have to bring this "not true" still two steps outward and this makes me recall my calculus lecturer Laszlo Czach who wrote the following sentence on the blackboard at the first day:

Every woman has a moment in her life when she'd like to do that's not alright.

After everybody stopped giggling he asked us to negate the line from the old song. And it is:
There is a woman who has no moment in her life when she'd like to do that is not alright.
Or taking the negation more inward it disappears:
There is a woman who at every moment of her life if would like to do something, it's alright.
In our case the negation should be moved outward not inward and we'll get:
For every ε it is not true that

for every M there is an N that $N > M$ and $|\frac{n}{N} - p| \geq \varepsilon$.

I wrote the "not true" part in a new line because we won't bring this not true further out.
Rather, we observe that this second line is a $P(\varepsilon, M, N)$ property if we regard p as fix.
Choosing one fix ε too, we get a property of M and N.
But here M is merely claiming the infinity of N. So what we have is the claim that the $Q_\varepsilon(N)$

$= |\frac{n}{N} - p| \geq \varepsilon$ property is occurring infinite many times.

I placed the ε as subscript because actually we have a different property for every ε value.
The point is that this property must occur infinitely and so our claim, the negative of this means that $Q_\varepsilon(N)$ must be finite many that is stops.

So what we expect is not a tendency of the future at all, rather a stopping of the past.
Of course, we can regard the negative, that is the continuing of Q_ε in a sequence too but then this will be a property of the non random sequences that is a strangeness.
You might realize that the whole argument could be applied for any approach of 0 or so called diminishing. Indeed, the Law Of Large Numbers simply says that the success ratio's difference from the p chance value must diminish. But diminishing means that the values stay under any desired ε value. Which with altered negatives means that they stop exceeding any chosen ε value. So why didn't I attack the name "diminishing" which is just as stupid because it falsely reflects an infinite tendency. Well, I already did even more deeply when in the previous section introduced it as being concentrated.
The Law Of Large Numbers is not enough to define random sequences. To be historically faithful we have to show how it received its death sentence from within the pack already.
Because a seemingly even more mystical expectability, the Law Of Dual Oscillations refuted it.
Von Mises' basic idea was to extend the Law Of Large Numbers to every observational sequence that can be obtained from a random s. Indeed, if we for example only watch every second outcomes then in this subset we must have the same success ratio limit as in the whole.

This grand generalization of the Law Of Large Numbers was supposed to guarantee all expectabilities.
In the coin toss sequences the half half probabilities mean that both the heads and tails success ratios approach half as limit. So the heads and tails equalize in this ratio sense.
Which of course doesn't mean that the actual numbers of heads and tails would get close.
In fact, we should expect bigger and bigger differences in their numbers by the following logic:
We must have infinite many perfect equalizations and after these regarding arbitrary long windows we must have all heads or all tails too.
The used assumption of infinite many equalizations follows from a much simpler intuition.
Namely, that the heads or tails can not remain for ever less than the others. They have to overcome the others in both directions and this is the earlier mentioned Law Of Dual Oscillations. Then indeed, if we had only finite many equalizations then from the last we would have more or less heads than tails for ever.
By the way, the arbitrary large excess of heads and tails also shows that the Law Of Large Numbers is not a simple plausibility claim. But in contrast the two directionality of the oscillations is a direct plausibility.
Observe also that mathematically we can have the success ratios of both the heads and the tails approach half and still defy this plausibility. We can have always more heads than tails or in reverse. Indeed, one simply would approach half from under while the other from above.
Ville realized that for any system of observational sequences we can similarly create a sequence that this one sidedness would happen in spite of all the observational limits approaching half.
This became well known very soon and so it started to sink in that maybe the Law Of Large Numbers is not the correct universal expectability.
To see that the Law Of Dual Oscillations is again merely a beginning stopping is quite easy.
Lets regard first the segment property that say the 1-s are more than the 0-s. This has half chance. Now lets regard the segment property that all sub beginnings as segments also have this property. This has much smaller chance. If in an even long segment we had this property then actually the 1-s have to be more than the 0-s by at least 2 while in odd segment they can be more by simply 1 too. So an even long segment can be continued by both 0 or 1 and we inherit the property while an odd can not be continued by 0 if it had only 1 excess of the 1-s over the 0-s. So these 0 continuation exclusions are the number of these minimal difference situations. By the Law Of Large Numbers these are actually becoming the main percentage of all cases. And so we will cut out almost all percentage of the cases at the these even long segments. So we do get a diminishing chance for the 1-s to be more than the 0-s. In fact, with a closer look it turns out too that the sum of these chances is finite and so this is the cause of the stopping. So the oscillation duality is not a continuing mystical expectability of the beginnings, rather simply the stopping of the non oscillation or majorizing, due to its fast diminishing in any increasing segments. And then we can apply this to beginnings too.
The truth is that we humans have no valid time plausibilities yet surfaced historically! That's why paradoxical false ones delude us. The most typical relating to randomness is the insane but instinctive belief that after some beginnings, certain outcomes are more or less likely.
A friend of mine Peter Elias showed me a very powerful Roulette strategy that plays on columns, that is on one third chances. We went with some other refugees to the Vienna Casino to try out the system and I said "Lets wait a few spins until a column is avoided for longest and start on that". Peter said "you are crazy". And I was. His father was a gambler and he calculated chances for long too, while I never bet in my whole life, so my "craziness" is forgivable. After this incident I tried to find the perfect didactical method that "enlightens" anybody about his false instinctive feeling. As it turned out it is simply the familiarity with the purely spatial character of the chances and outcome sequences. If someone writes down all the thirty-two possible five segments of a coin flip then he will "see" that inside these, the fourth outcome is dual for every earlier three. The feeling that after ten heads we must "already" get a tail, comes from the valid feeling that these are very rare. So we jump to the false conclusion that eleven heads must be much rarer than ten. But it is actually only twice as rare. And so to encounter a head after ten heads has exactly half chance just like encountering a head as start. Explaining it like this helps

but to truly overcome the false intuitions we simply have to identify the chances as ratios of successes to the number of all possibilities.
The best introduction to these "chance countings" is card pullings from a deck.
Pulling cards is not mere repetition because after a pull we have less cards to pull from.
So, pulling three cards has not 52 x 52 x 52 many cases only 52 x 51 x 50.
Similarly, to pull three aces has 4 x 3 x 2 = 24 many cases. So the chance of pulling three aces is 24 divided by 52 x 51 x 50.
But the counting of the desired cases can have some further tricks too.
For example, pulling at least one ace from three pulled cards first seems to be logical by regarding three situations. Namely pulling 1 or 2 or 3 aces. In the first case, the ace can be the first which has 4 possibilities and the other two pulls 48 many. This is 4 x 48 x 48 possibilities but similarly the ace can be the second or the third, so we have 3 x 4 x 48 x 48 possibilities only in this first situation. Then we have to add the two ace and three ace situations.
As we see this is quite complicated. A much easier way is to calculate the cases of the complement feature, which is not pulling any ace. This has simply 48 x 47 x 46 many cases and so 52 x 51 x 50 – 48 x 47 x 46 is the number of the successful cases for our original feature and dividing this with 52 x 51 x 50 is the chance.
A typical source of confusions is that we use chances in everyday situations when in fact we have no such meaningful cases. For example, the weather forecast may say that the chance of rain is 50 percent. Suppose it doesn't rain till noon. Then we can ask what is the chance of rain for the rest of the day. Some may say it's the same, some may say it's less because the forecast failed already for the morning and some may say it's more because the no rain till noon pushes the rain to the afternoon. In truth, we have no correct answer.
A famous case of false intuitions is the "Show Host Help" situation. We have a show where there is a prize behind one of three curtains. The contestant chooses but then before revealing our choice the show host offers a chance to alter. In fact, he reveals one of the not chosen curtains as empty. Should we switch to the other not revealed curtain or stick with our original choice? Many feel that it doesn't matter. Even some eminent mathematicians did so. In truth, a switch increases our chances from $\frac{1}{3}$ to $\frac{2}{3}$. Indeed, regard the possible games for a year.
The original choice obviously is correct in about one third of the 365 days. If we switch then all these days we'll loose the prize but in all other days we get the prize.

4. Beginnings And Sequences

b = beginning = any finite sequence of 0-s or 1-s.

s = sequence = any infinite sequence of 0-s or 1-s.

S = set = any set of sequences

a c beginning continues b if b is sub beginning of c.
an s sequence continues b if b is a beginning of s.
b being a beginning of s or s continuing b is also expressed as " b covers s ".
The reason for this expression is that the beginnings can be visualized as halving intervals while the sequences as points on the [0 , 1] interval. For example, 001011 is a sixth level halving interval. It is one of the $2^6 = 64$ equal ones and is located at the left left right left right right choices in the successive halvings of [0 , 1]. On the other hand the 00101110001011 . . . sequence is an actual point that the successive halvings approach. And this is indeed covered by the previous interval because it was a beginning of it. So this vision is like an infinite road that has as pathways the sequences and as final target the [0 , 1] interval. Since it was Kolmogorov who introduced this, I call it the Kolmogorov Road.

B = beginning pool = pool in short = any set of beginnings.

B covers s if some b in B covers s so simply if s has a beginning in B.
This means that we regard B as the "or" collection of beginnings as far as covering goes.
B covers S if B covers every s element of S.
The covered sequences as set by a b or a B is denoted as [b] or [B].
In short we just say that this is the set covered by b or B.
There is a crucial way to refine the covering relation of a pool and a sequence or a set. Namely:
s continues in B if there are infinite many beginnings of s in B.
S continues in B if all s elements of S continue in B.
Visually these mean that the s pathway or S pathways can be traveled in B.

$B_1 , B_2 , \ldots$ = pool sequence
The $B_1, B_2, ..$ pool sequence covers s or S if every B_n covers s or S.
In other words s has a beginning in each B_n or every s in S has a beginning in each B_n.
So a pool sequence as coverer is a mixture of "or" and "and". Indeed, the pools themselves are an "or" meaning but the sequence members are an "and" meaning.
So it might seem that this covering is a much stronger prediction feature than continuing because here we tell exactly infinite many choice pools for the beginnings while at continuing we only require to have infinite many from a single pool.
But in some sense a covering is weaker. Indeed, having a beginning from each pool of a sequence can mean only to have finite many beginnings chosen if some are chosen infinite many times. For example being covered by a single pool repeated as sequence easily allows this.
In fact even with a single beginning in it. To exclude these obvious failures is easy as follows:

minB = the minimal length with which there is b in B.
$B_1, B_2, ..$ is long if $\min B_n$ becomes arbitrary big for some n member.

If $B_1, B_2, ..$ is long then any s that is covered by $B_1, B_2, ..$ will continue in
$B_1 \cup B_2 \cup \ldots$
Indeed, longer and longer beginnings of s will be covered and so they continue.
And now the bigger surprise:
For any B pool there is a $B_1, B_2, ..$ pool sequence that:
s is covered by $B_1, B_2, ..$ $\leftrightarrow$ s continues in B
To prove this we need a useful concept, the length groups of a B pool:

B^n = the n long beginnings appearing in B. So it's a as a sub pool of B

The most logical application of this is that any B can be partitioned into its length groups as:
$B = B^1 \cup B^2 \cup B^3 \cup \ldots$
But these wouldn't be good as a covering sequence because a continuing s can skip members.
We need a different application, namely a shrinking sequence by subtracting the length groups.
So, $B_1 = B$, then $B_2 = B - B^1$ that is we leave out the single 0 or 1 if they appeared in B.
Then $B_3 = B - B^1 - B^2$ so we leave out the two length ones too. And so on.
Obviously this sequence is long again and also B_1 is the union and so our previous result implies the $\rightarrow$ direction at once. For the $\leftarrow$ direction observe that any beginning appearing in B_n automatically appears in the earliers. So enough to show that arbitrary late members contain beginning of s. Which is trivial because all n long beginnings from $B = B_1$ are still in B_n and s has arbitrary long beginnings in $B = B_1$.

This trivial construction might suggest that this whole covering idea was useless but actually it was the original approach to the Law Of Stopping. So we give now an alternative way of replacing a single B pool with a covering pool sequence:
A pool is "excluding" if no elements in it are continuing each other.
Indeed, such beginnings exclude each other as single trials.

a b is minimal element in B if it is element and has no sub beginning in B.

B^0 = the minimal pool of B = the pool of all the minimal elements of B.

B^0 is an excluding sub pool of B and they cover the same sequences: $[B] = [B^0]$.

But there can be other sub pools of B that satisfy both of these conditions.
B^0 has a third advantage too, namely that every element outside, that is in $B - B^0$ has to be a continuation of some element in B^0. So then:
$B^{(1)}$ = the 1 sub beginning sub pool of B = $(B - B^0)^0$.
Indeed, the beginnings in this set have exactly one sub beginnings, namely in B^0.
$B^{(2)}$ = the 2 sub beginning sub pool of B = $(B - B^0 - B^{(1)})^0$. And so on.
And now a new "perfect" partition of a B is:

$B = B^0 \cup B^{(1)} \cup B^{(2)} \cup \ldots$
These now work at once as a covering sequence because the beginnings can not skip them.
And these are again a long sequence.

$\langle b \rangle$ = b's length = how many digits are in it.

$| b | =$ b's chance value $= \dfrac{1}{2^{\langle b \rangle}}$

$\langle B \rangle$ = B's size = B's chance total $= \sum_{b \in B} | b |$

$|B^n|$ = the chance of $B^n = \dfrac{\text{num}\, B^n}{2^n}$ where num B^n is the number of elements in B^n.
Remember that B^n is a set of n long beginnings. So indeed we can regard all possible n long beginnings as all possible cases and the ones in B^n as favorable ones giving $|B^n|$ as chance.
This was true because the cases were excluding and had equal chances.
Of course $|B^n| = \langle B^n \rangle$ which reflects the law that excluding chances are additive for "or".
We'll come back to more general B pools where the chance total has a chance meaning.
Namely, the excluding pools. But right now we introduce a crucial distinction for the general $\langle B \rangle$ sizes too that have no chance meaning:

B is narrow if $\langle B \rangle$ is finite.
Usually B is an infinite set, so we have infinite many $\dfrac{1}{2^n}$ fractions added together to get $\langle B \rangle$.
Also, one value can even appear more times because same long beginnings in B have same chances. This fact that infinite many numbers can sum up to be finite, could be called the Sum Limit Paradox and the Greek mathematicians struggled with this first. Luckily for us, the infinite decimals are such infinite sums of similar fractions too but with ten, hundred and so on denominators. The numerators are the digits themselves:
$3.14159\ldots = 3 + \dfrac{1}{10} + \dfrac{4}{100} + \dfrac{1}{1000} + \dfrac{5}{10000} + \dfrac{9}{100000} + \ldots$
So we have this "silver platter", that the Greek mathematicians didn't know at all.

Still, the simplest infinite sum is: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \; . \; . \; . \; = 1.$

If we start from a meter from a wall and go forward a half meter then we become half meter away because $1 - \frac{1}{2} = \frac{1}{2}$. Then going a quarter we become a quarter away because $\frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ and so on. So indeed, we approach the wall by these steps exactly. Algebraically, our argument was a tricky replacement of the members because we knew the sum as start. Namely:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \; . \; . = (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{4}) + (\frac{1}{4} - \frac{1}{8}) + \; . \; . = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{4} + \frac{1}{4} - \frac{1}{8} + \; . \; . = 1$$

A more precise proof uses this same trick but for a finite claim, namely as :

$$\frac{1}{2} + \frac{1}{4} + \; . \; . \; . + \frac{1}{2^n} = 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{4} + \; . \; . \; . + \frac{1}{2^{n-1}} - \frac{1}{2^n} = 1 - \frac{1}{2^n}$$

This then really means that the finite partial sums of the infinite sum approach 1.

The above definition of B being narrow might seem as useless if we jump to the false conclusion that such smaller and smaller values always add up to finite value anyway.

So the truth that smaller and smaller values can be infinite, could be called the Counter Sum Limit Paradox. To see why this is true, we should imagine first infinite many fix values say 1 added. This is obviously infinite and now we should cut these into more and more parts and thus obviously get smaller and smaller values:

$$1 + 1 + 1 + 1 + \; . \; . \; . = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \; . \; . \; . \; = \infty$$

Of course, we could alter these members by increasing a bit some and decreasing the others and so we could get strictly decreasing values too.

Amazingly, the single simple fractions already add up to infinity:

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \; . \; . \; . \; = \infty$$

The proof of this is again to start with infinite many same values but these should be now all $\frac{1}{2}$.

Then we again replace these with new members and then we even increase some members:

$$\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \; . \; . \; . \; = \infty .$$

$$\frac{1}{2} + \overbrace{\frac{1}{4} + \frac{1}{4}} + \overbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}} + \overbrace{\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \; . \; . \; . + \frac{1}{16}} + \; . \; . \; . \; = \infty .$$

$$\wedge \qquad \wedge \quad \wedge \quad \wedge \qquad \wedge \quad \wedge \quad \wedge$$

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{11} + \; . \; . \; . + \frac{1}{16} + \; . \; . \; . \; = \infty .$$

The finite sum cases are of course much more interesting because then to find out the actual total is a big task itself. It took Euler a year to prove the following fact:

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \; . \; . \; . \; = \frac{\pi^2}{6}$$

Some positive $v_1, v_2, \; . \; . \; .$ values are diminishing or approach zero or in symbol $v_n \to 0$ if :
for every ε positive value there is an N number, so that if $n > N$, then $v_n < \varepsilon$.

So the v_n values not merely get arbitrary small rather they stay arbitrary small after a point.

Some positive $v_1, v_2, \; . \; . \; .$ values approach infinity or in symbol $v_n \to \infty$ if :
for every M natural there is an N , so that if $n > N$, then $v_n > M$.

Again, this means that the values not merely get arbitrary big rather they stay arbitrary big.

The claim that they just get arbitrary big is also called being unbounded.

Remember that $B_1, B_2, \; . \; .$ being long meant that $\min B_n$ is unbounded.

$B_1, B_2, \; . \; .$ is called diminishing if $<B_1>, <B_2>, \; . \; .$ is diminishing.

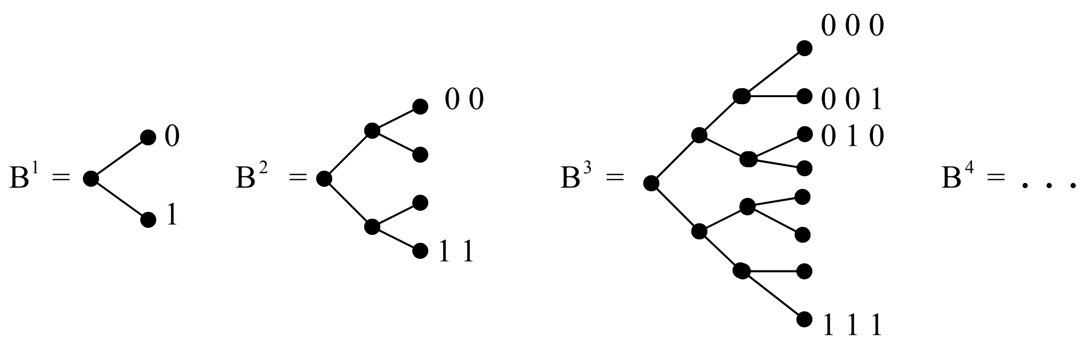
A good exercise to see these limits in action is to prove that:
$B_1, B_2, ..$ is diminishing $\rightarrow$ $B_1, B_2, ..$ is long
Obviously it's enough to prove that:
$<B_1>, <B_2>, . . . \rightarrow 0 \quad \Rightarrow \quad \min B_1 , \min B_2 , . . . \rightarrow \infty$
The negative of the second means that $\min B_n$ is under a fix M value infinite many times.
This means infinite many beginnings that are all shorter than M.
These all have at least $\frac{1}{2^M}$ chance value and thus can not stay under an $\varepsilon < \frac{1}{2^M}$.

5. Nil Sets

S is nil set if for every ε there is a B pool that covers S and $<B> \leq \varepsilon$.
S is zero set if for every ε there is a finite B pool that covers S and $<B> \leq \varepsilon$.
The real intuitive vision of nil and zero sets comes through the Kolmogorov Road:
On an interval an S point set is nil if for any small ε there is a set of intervals $\{I_1 , I_2 , . . . \}$ so that these cover S and their total length is maximum ε that is $|I_1| + |I_2| + \, . . \leq \varepsilon$.
For zero point sets the covering interval sets are finite.
If an S set is sequencable as $P_1 , P_2 , . . .$ then S is a nil set because we can cover P_1 with an $\frac{\varepsilon}{2}$ long I_1 also P_2 with an $\frac{\varepsilon}{4}$ long I_2 , and so on and the total is $\frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \, . . = \varepsilon$.
This is the most visual way of seeing that a sequence can not exhaust a full interval because then an arbitrary small ε could cover the full interval. In fact, we feel that an interval should not be coverable by any set of intervals that totals less than the interval's length.
To prove this precisely is quite involved and so this seemingly easiest way of seeing that an interval is not sequencable is actually the most complicated. In fact, this road is usually not mentioned at all, rather Cantor's anti diagonal method is used for the decimals or Cantor's common point axiom is used to show a point outside a point sequence.
The anti diagonal method simply shows that for any sequence of infinite decimals we can create a decimal definitely missing from our sequence. Indeed, we can alter all digits in the diagonal that is the first digit from the first the second from the second and so on and this altered diagonal decimal can not be any of the members in our sequence.
Cantor's other important realization was that the rationals, that is the fractions are a mere sequence. This is again easy to see if we list the fractions by increasing totals of the numerator and denominator. One fix total can have only finite many variants. This grand idea of going in finite groups can show many other seemingly big sets to become merely a sequence.
The sequencability of the fractions goes against how they are located on the number line.
Indeed, they are "dense", meaning that every interval contains a fraction.
Which of course means that every interval must contain infinite many.
The irrationals might even feel as merely the "holes" among the fractions and they are also dense.
We don't see the big difference between these two sets and why the irrationals must be so much more than the rationals. And this "more" means a set that is not sequencable but also a set that is not coverable by smaller total than the full interval. These two comparing of sets by equivalence or by coverability is what we call cardinality and measure.
For the Greeks even the existence of the irrationals was a problem. With the silver platter of the infinite decimals the situation is much clearer if we also use the simple fact that fractions are always decimals periodic from a point. This is a trivial consequence of the digital division process taught in elementary schools. Indeed, we can only have a finite choice of remainders.
So, all the non periodic decimals are irrationals and these are plenty more. Cantor made this subjective truth objective and thus created a new proof for the existence of irrationals without looking at the decimals. His method was then usable for wider special sets of numbers, namely for the algebraic ones, where a decimal method was already known too but was much more complicated than the periodicalness of the rationals. To go all the way would have meant to say that all the effective decimals are still just a sequence. Indeed, for every given n number we have

only a finite N number of possible systems using exactly n many basic parts in a given framework. So we can list all possible effectivities of a framework.
Cantor didn't go this way but did go deep into cardinality versus measure
If an S set has smaller measure than T then the cardinality of S can not be bigger than T's.
But they can have same cardinality and the simplest paradox of the points shows this.
This is the projection of a smaller interval onto a bigger. The measure size increases but the cardinality is the same. The paradox is not this fact, rather the visually trivial cardinality equality, that is equivalence. Indeed, a projection orders to every point a unique one and so it seems that points are squeezable. But they also seem like fix objects.
It's almost a mathematical analogue of the particle wave duality of physics.
An other even more drastic confrontation of measure and cardinality is the fact that already the smallest measure, that is being a nil set allows a set to be equivalent to the full interval.
This shows that the reversal of our result that all sequencable sets are nil sets, is not true.
This example was also Cantor's discovery, in fact it is called the Cantor set and actually it is not just nil set but zero set. It is obtained from the unit interval by carving out the middle third interval, then again carving out the middle third from the two remaining ones, and so on.
The carvings leave $\frac{2}{3}$ portion at every step and so the Cantor set formally has measure :
$1 \text{ x } \frac{2}{3} \text{ x } \frac{2}{3} \text{ x } \ldots = 0$. More precisely, after the successive carvings we have remaining intervals having total lengths: $\frac{2}{3}$, $\frac{2}{3} \text{x} \frac{2}{3}$, $\frac{2}{3} \text{x} \frac{2}{3} \text{x} \frac{2}{3}$, ... and these diminish. The carved out total is harder to calculate as: $\frac{1}{3} + (1 - \frac{1}{3}) \text{x} \frac{1}{3} + (1 - \frac{1}{3} - (1 - \frac{1}{3}) \text{x} \frac{1}{3}) \text{ x } \frac{1}{3} + \ldots$
The exact third middle carvings were used by Cantor merely to be specific because the main point back then was that the remaining zero set is equivalent to the full interval.
This follows from the fact that the remaining set is also a sequence of possible dual choices.
Using infinite binary numbers, that is repeated halving intervals, we can also carve out middles even if they are not exactly at the middle and we can make the carved out total anything up to 1.
Zero sets have a perfect characterization as follows:
An inner point of a set contains a full surrounding in the set. The set of all inner points is the interior and this on a line forms non overlapping intervals. Their total length is the interior size.
A set in a unit interval is zero set if and only if the complement has interior size 1.
Outside an ε long subinterval of the unit interval the interior total can maximum be $1 - \varepsilon$ and so a zero set's complement must have inner point in every subinterval.
This feature of the zero sets can be expressed in a second more visual way too:
A point of the space that has a surrounding fully outside a set is an outer point.
And if we use "everywhere" to mean in every interval then the zero sets are everywhere outer.
They have outer point in every interval. As a third way we can use the already mentioned density:
A point set is dense in an interval if every sub interval contains point.
This of course means that there is no outer point in the interval.
So being everywhere outer could also be called as being "nowhere dense".
The reversal of the fact that zero sets are everywhere outer is not true trivially, because we can carve out less than 1 in total and so we get everywhere outer set with positive measure.
But more surprisingly, the reversal is not true even by assuming that the everywhere outer set is nil set. Indeed, we can have a sequence of carving intervals with total under 1. But then instead of carving them out we just regard their end points. This set is a sequence so is a nil set and the only inner points of the complement are the points inside the intervals. So the interior size is under 1.
The difference between nil sets and zero sets is even more striking if we return to sequences.
First of all, here density means that for any b beginning there is an $s \in S$ with b being beginning of s. A much stronger property is if S is beginning independent, meaning that for any $s \in S$ we can alter any beginning of s to any b and this altered sequence remains in S. This has no visual relevance in an interval and yet we all feel that the random sequences as a set must

be beginning independent. A strangeness doesn't have to be beginning independent but amazingly a perfect strangeness criteria for randomness does have to be.
More precisely, such criteria is a collection of strangenesses and their "or" collection is then a single universal strangeness which is the complement of randomness. And it is trivial that if a set is beginning independent then its complement is too. So a universal strangeness has to be a beginning independent nil set. But not all beginning independent nil sets are universal.
For example, having all 0-s from a point in an s sequence is a non universal strangeness.
This as S set contains exactly the sequences that continue all possible beginnings with 0-s.
So obviously S is beginning independent but it is not a universal strangeness because not being such all 0 ending sequence can be strange in many other ways.
By the way, the beginnings can be listed as:
{ 0 , 1 , 00 , 01 , 10 , 11 , 000 , 001 , 010 , 011 , 100 , 101 , 110 , 111 , 0000 , . . . }
So S itself is merely a sequence and thus it is a nil set as it should be by being a strangeness.
Obviously, an S set of sequences is nil set if and only if there is a $B_1, B_2, ..$ diminishing pool sequence that covers S. But this still involves infinite many pools.
For zero sets all the B_n members have only finite many members and for such pool sequences, the covering can be replaced by using only a single pool.
Namely, we can introduce a $\sim B_n$ complement of each B_n member so that an s is in $[B_n]$ exactly if s is outside of $[\sim B_n]$. All we have to do is continue every member of B_n to the longest ones and then regard the missing ones among these in the single length group.
Then S is covered by $B_1, B_2, ..$ if and only if $\neg S$ is covered by $\sim B_1 \cup \sim B_2 \cup ..$
Where $\neg S$ is the complement set of S, that is the set of all sequences that are not in S.
Finally, we can regard the minimal subset of the union above and so we get the result that corresponds to the already mentioned fact among point sets that zero sets are the complements of non overlapping intervals that total 1. In other words:
Kurtz's Raw Criteria:
S is a zero set $\leftrightarrow$ $\neg S$ is covered by an excluding B that $\langle B \rangle = 1$.
So here among sequences, a zero set is simply the remaining sequences in the binary tree:



if we block out so many beginnings that have a total of 1 in chance value.
To be more specific, we go through each group above and block out nothing or some members.
When we go to the next group we only look at the continuations of yet non blocked paths and we may block out new beginnings again. We simply prune the binary tree. The pruned beginnings are an excluding set of beginnings and we also assume that they grow to 1 in chance total.
If we would reach the 1 value already at a group then as we'll show it soon, we would have no paths to continue. But amazingly, if we only approach 1 then quite a lot of paths can remain.
Indeed, the Cantor set proved that the left possible sequences can be same many as all possible sequences. Meaning this in the equivalence or cardinality sense of course, and this has not much relevance for us. But it does suggest that these zero sets of sequences could also be used together as universal strangeness. Individually they are always nowhere dense but together they are beginning independent and so dense too. This was exactly the already mentioned strangely late realization of Kurtz to use the Law Of Occurrence as a randomness criteria.

The real advantage of this is that every zero set can be defined by a single pool, namely by the pruned beginnings that cover the complement.
Now we want a similar single pool replacement for nil sets in general.
Remember that we had a result that replaced a covering by a pool sequence with continuing in a single pool. In fact, I showed two constructions for the $\leftarrow$ direction.
Remember also that we defined narrow pools as having finite chance total. Using these two:
Solovay's Raw Criteria:
S is a nil set $\leftrightarrow$ S continues in a narrow pool.
The $\rightarrow$ direction seems easy! Indeed, diminishing cover is long and so our earlier result shows that a covering implies continuing in $B_1 \cup B_2 \cup \ldots$
Unfortunately, we don't see why this combined set would be finite totaled. And it isn't.
What we have to do, is to regard a $C_1, C_2, ..$ subsequence of $B_1, B_2, ..$ where:
$< C_1 > \le \frac{1}{2}$, $< C_2 > \le \frac{1}{4}$, $< C_3 > \le \frac{1}{8}$, and so on.
This again covers all s so the previous argument applies for this subsequence but it is now maximum $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$.
The $\leftarrow$ direction is easy by our earlier first construction of subtracting the longer and longer length groups. The chance totals of these approach $< B >$ and so the diminishing is trivial.
We will need later the other construction too, that is the perfect distribution members.
To see that these diminish is much harder. It has members that are all excluding pools.
First we prove that such excluding E pools can have only a chance total of maximum 1.
We show that the finite beginning sum is less than 1.
In the $E^1 \cup E^2 \cup .. \cup E^n \cup \ldots$ length distribution of E we replace all beginnings in the members up to n with beginnings of n lengths.
To have the same total is guaranteed by simply to continue them in all possible manner up to n lengths. Indeed, every single continuation then doubles the number of them but also halves the chance values. And of course, the chance total of all possible n long beginnings is 1.
But we won't get this as total because there have to be some n long beginnings missing.
Indeed, there are still longer beginnings in E that can not be continuations of neither original n long ones nor artificially continued ones. The only problem would be if our artificial continuations would coincide with already existing ones or each other because then this total would be false. But this is impossible due to the non continuingness of E again.
Unfortunately, this fact that the members in $< B > = < B^0 > + < B^{(1)} > + < B^{(2)} > + \ldots$ are all maximum 1 , doesn't mean anything about them being diminishing.
Instead, we'll need that they can not increase : $< B^0 > \ge < B^{(1)} > \ge < B^{(2)} > \ge \ldots$
But even if we had definite decreasing values here, it still wouldn't mean that they diminish.
Indeed, they could decrease and approach a positive value. So we'll need more than this.
But as start, our claim is that $< B > \ge < E >$ if all E elements continue some B element but no two elements of E continue each other.
Lets regard all those $e_1, e_2, \ldots$ in E that continue a common b of B.
Lets cut off the common b from them and regard $e_1 - b$, $e_2 - b$, ..
The chance value of any e is the product of b's and (e – b)'s chance. So the total chance of the e-s is simply the chance of b multiplied by the total chances of (e – b)-s.
These can not continue each other because the e-s weren't either. So by what we proved above, the total of the (e – b)-s chance is maximum 1 and so the total of the e-s chance is maximum the chance of b. And since all elements of E are such continuations, the total chance of E is maximum that of B's too.
Using what we just proved for the first n + 1 members, we get :
$(n+1) < B^{(n)} > \le < B^0 > + < B^{(1)} > + < B^{(2)} > + \ldots + < B^{(n)} > < < B >$
So $< B^{(n)} > < \frac{< B >}{n+1}$ and so the members indeed diminish.

And finally, if s is continuing in B then s is covered by B^0, $B^{(1)}$, $B^{(2)}$, . . because every beginning in $B^{(n)}$ brings in beginnings in all the earliers too. Indeed having n sub beginnings in B implies to have sub beginning with any smaller number of sub beginnings in B.
The finally still used fact that for excluding E we have $< E > \leq 1$ suggests that $< E >$ should be abbreviated as $| E |$ because it can be regarded as a chance.
And indeed, this $< E > = | E |$ value gives the chance that a randomly chosen infinite sequence will have a beginning from E. This E is the same pruned beginnings that we already used with 1 total. To see the general meaning of $< E > = | E |$ is again easiest by envisioning E as prunings, that is by the increasing length groups as :

$\{\{b_1 , b_2 , . . , b_k\} \cup \{ \quad \} \cup \{ \quad \} \cup \ . . . \}$.

The total of b_1 , b_2, . . . ,b_k-s chance is $\frac{k}{2^n}$ where n is their length and this is exactly the chance that such minimal long beginning from E will occur in a randomly chosen infinite sequence. Then the next equal ones' total gives the chance that that long will occur, and so on. And these groups are independent because we don't have continuations and so the full chance total of E gives the chance of "or" , that is the chance of any occurrence from E.
The paradox is the already mentioned fact that 1 total chance can leave out a lot of sequences.
First of all, if E is finite then there is no such paradox. Any possible b beginning is either a sub beginning of one in E or has a sub beginning in E , in other words continues one in E.
Indeed, we can extend again all members of E to the length of the longest ones with all possible continuations and we end up with 1 total in a set E* of all same n long beginnings.
So here in E* all n long beginnings must appear. Thus, for this set the claim is trivial because any shorter than n beginning is sub beginning and any longer is a continuation. Also, this inherits to E. For the longer than n beginnings trivially and for the shorter ones because any such is a beginning in an extension and so either it was beginning already in the extended one or continues it. What we showed equivalently means that any sequence is covered by our E set.
But the same is not necessarily true for an infinite E.

For example let be $E = \{ 0 , 10 , 110 , 1110 , 11110 , \ . \ . \ . \ \}$.
It is excluding, $< E > = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ . \ . \ . = 1$ and yet the sequence 111 . . . is not covered by E because the 1 , 11 , 111 , . . . beginnings are neither sub beginnings nor having beginnings in E.
This example gives the impression that the uncovered sequences are only a few but as we explained earlier for the same pruning sets, the left out zero set of sequences can be huge in cardinality. Still these are strange, that is a random sequence must have a beginning from E.
This is the Law Of Occurrence.

6. Quota

The concept of cover can be generalized from covering sequences to pools covering each other. Namely, through [B] , the set of covered sequences, by B. So C covers B if [C] $\supseteq$ [B].
Then $C_1, C_2, ..$ covers $B_1, B_2, ..$ if every C_n covers B_n .
B^0 is a minimal cover of B. So it covers B and any C cover of B covers B^0 too.
The sizes don't follow the coverings, so a covering pool can have smaller size than the covered. For example {0 , 1} covers the pool of all beginnings. This size overgrowth can not happen if the covered set is excluding. As a special case if C covers B then $< C > \geq < B^0 >$.
We might have a covering pool sequence for an S set without being able to tell the minimal sets of the members. Luckily there is a practical inequality for the size of the minimal sub pool using a more natural concept the groups.
$B = B^1 \cup B^2 \cup B^3 \cup \ . \ . \ .$ which implies $< B > = <B^1> + <B^2> + <B^3> + \ . \ . \ .$ and the members in this sum could be called the portions of B. This name reflects not just that these members are portions of the total but that they themselves are portions from all possible fix long beginnings. If $q \geq < B^n >$ then this q could be called as a portion bound for B.
Of course, q = 1 is a trivial portion bound for any B.
The lowest portion bound is denoted as q (B), and will be called the quota of B. We could call this the maximal group portion of B too, but since we have infinite many groups, it doesn't have to be an actual portion value in any of them. It may only be the limit of them.
The first fact we show is that if C is a cover of B then $< C > \geq q(B)$.
It's enough to show that every group's portion in B can't be more than < C >.
Those b elements in the group that are covered by same or earlier group elements of C, are okay because these early elements of C have a total chance value at least as big as the total of these b-s. The remaining b elements in the group, that are covered only by later elements of C, can be replaced first by { b0 , b1}, then by { b00 , b01 , b10 , b11} and so on.
These all total the same in chance as b and all these elements must be covered by C elements gradually. C might even cover some branches more times, but just regarding the first ones, we clearly have at least as big total in C as these b elements amounted to.

Now we view the $B_1, B_2, ..$ pool sequences so that every B_n member in it, is partitioned into its $B_n^1 \cup B_n^2 \cup \ . \ . \ .$ groups. Thus the quotas of the members can be regarded too.

As n increases, we want $q(B_n) \to 0$. First of all, this implies being long at once.
Indeed, if $q(B_n)$ is less than $\frac{1}{2^m}$, then up to m, we can't have beginnings at all in B_n .
Even though $q(B_n) \to 0$ is much stronger than being long, it is also trivially weaker than the $<B_n> \to 0$ diminishing. Indeed, this implies it at once, since:
$< B > = < B^1 > + < B^2 > + \ . \ . \ . \ > q(B)$.
But by $< C > \geq q(B)$ we also have that any $C_1, C_2, ..$ that covers $B_1, B_2, ..$ implies the same that is $<C_n> \to 0$ implies $q(B_n) \to 0$.
To see that the reverse is not true, I show two interesting pool sequences. In both we'll have $q(B_n) = \frac{1}{2^n}$ but no cover of B_n can be smaller than 1 because B_n covers all sequences.

I list the same grouped members under each other and marking their common sub beginning:

B_1 = { 0 , 10	B_2 = { $\overline{00}$,	$\overline{01}$0 ,	$\overline{10}$00 ,	$\overline{11}$000
11 }		011 ,	1001 ,	11001
			1010 ,	11010
			1011 ,	11011
				11100
				11101
				11110
				11111 }

B_3 = { $\overline{000}$,	$\overline{001}$0 ,	$\overline{010}$00 ,	$\overline{011}$000 ,	$\overline{100}$0000 ,	$\overline{101}$00000 ,	$\overline{110}$ . . .
	0011 ,	01001 ,	011001 ,	. . .		
		01010 ,	011010 ,	. . .		
		01011 ,	011011 ,	. . .		
			011100 ,	. . .		
			011101 ,	. . .		
			011110 ,	. . .		
			011111 ,	. . .		

The second example will be even more amazing because $B_1 \supset B_2 \supset B_3 \supset$ . . . also stands:

B_1		B_2 starts here				B_3 starts here
0 ,	10 ,	$\overline{00}$0 ,	$\overline{01}$00 ,	$\overline{10}$000 ,	$\overline{11}$0000 ,	$\overline{000}$0000
	11	001 ,	0101 ,	10001 ,	110001 ,	0000001
			0110 ,	10010 ,	110010 ,	0000010
			0111 ,	10011	.	.
				10100	.	.
				10101	.	.
				10110	16	32
				10111	elements	elements

$\overline{001}$00000 ,	$\overline{010}$000000 ,	$\overline{011}$0000000 ,	$\overline{100}$00000000
00100001 ,	010000001 ,	0110000001 ,	10000000001
00100010 ,	010000010 ,	0110000010 ,	10000000010
.	.	.	.
.	.	.	.
.	.	.	.
64 elements	128 elements	256 elements	512 elements

			B_4 starts here
$\overline{101}$000000 ,	$\overline{110}$00000000 ,	$\overline{111}$0000000000	$\overline{0000}$. . . .
101000001 ,	11000000001 ,	1110000000001	
101000010 ,	11000000010 ,	1110000000010	
.	.	.	
.	.	.	
.	.	.	
1024 elements	2048 elements	4096 elements	

In the first example there were no continuations in B_n but in the second we had to allow this to keep the quota. So the natural idea toward getting the cover lowered by the quota is to assume that B contains all continuations of its elements.
Indeed, for such "complete" B pool we have $< B^0 > = q(B)$.
And this is almost trivial because the B^0 elements' continuations will create same portions in all groups. So if for the $B_1, B_2, ..$ sequence of complete pools we have $q(B) \to 0$ then any S covered by $B_1, B_2, ..$ must be nil set too. This is indeed more general because these B_n members don't have to be narrow pools. In fact they can not be because a complete pool has infinite chance total.

7. The Betting Function

Ville was the one who destroyed Von Mises' false status quo of the Law Of large Numbers.
But Ville made a much simpler yet much more heuristic positive discovery too, that strangely remained unrecognized for decades.
If we asked someone how it could be proven that a sequence is not random rather altered or "fixed" in any sense, then sooner or later we would get the answer that such strangeness must be usable to make money. The casinos survive because there are little fix gains for them like the 0 number on a roulette. If this weren't then the house had only its fair chance like all players.
The rewards after bets are all very logical and have no built in gains for the house. The simplest is of course that if something has a half chance then we should get a double as reward.
Most gamblers play by deciding when to double. A more refined method is to gamble on both alternatives with different bets. Even the non betting "sober" moments can be regarded as bettings. Indeed, not to bet can be replaced by betting on both outcomes equally. Then one of the bets is lost, the other doubles, so we are back to what we had.
So we can regard gambling on dual half half chanced sequences by a β betting function.
$\beta(0)$ and $\beta(1)$ are given as starting bets. Either one wins we get double back so $2\beta(0)$ or $2\beta(1)$ will be our new total and we must spend all these on betting again but with our free distribution on the possible outcomes. So:
$2\beta(0)$ = new bet on 0 + new bet on 1 or $2\beta(1)$ = new bet on 0 + new bet on 1.
Ville's ingenious idea was to incorporate both of the new betting possibilities into β by regarding the new next outcomes as last members of a beginning.
So instead of the "or" of new outcomes, we'll have an "and" of beginnings.
$2\beta(0)$ = new bet at 00 + new bet at 01 and $2\beta(1)$ = new bet at 10 + new bet at 11
So now β is defined on beginnings and has as value the bets on the last elements:
$2\beta(0) = \beta(00) + \beta(01)$ and $2\beta(1) = \beta(10) + \beta(11)$
So, β is defined on b but $\beta(b)$ is not a bet on b rather on the last member of b.
$2\beta(b)$ is the win at the last member of b which is also our acquired total after b as outcomes.
This must be spent on bets on b0 and b1 and so : $2\beta(b) = \beta(b0) + \beta(b1)$.
A β betting function succeeds on an s sequence if $\beta(b) \to \infty$ while going through all b beginnings of s. Remember that this approaching infinity means not merely becoming arbitrary big, that is being unbounded rather staying above arbitrary value. And indeed, being unbounded that is achieving arbitrary big gains but always falling back under some fix value in our total is not a success in an infinite game.
To have a β that succeeds on a single s is very easy! We simply place all bets on the outcomes that are in s and place 0 everywhere else. At every step we'll double our money.
The interesting question is what S sets can be such that a β succeeds on all elements of S.
To get the largest possible S we can't use such primitive all on one, nothing on the other betting function but for a partial success we still "almost" can. This partial problem is to get a β only up to a given n length so that it will succeed on some given members of the n long beginnings.

The “almost” then means that all bets are either such all on one bets or halved equal bets, which as I explained practically means no bets.
Indeed, here all those beginnings that we don’t want, have a longest sub beginning after which all continuations are bad and so these must be simply avoided that is have zero bets. An extreme situation is for example if we want all n long beginnings that end with a 1. Then we don’t have such avoidable sub beginnings up to the last $n - 1$ length and so only here shall we actually gamble by simply bet on the desired ones all our moneys. The gains will be merely a doubling of our original money. And indeed, that’s the best we can do. The wanted set was simply such difficult. If it has earlier cut out continuations then of course we will get more gains on the good continuations. The minimal total covering of the wanted beginnings is a given portion and the maximal total gain is always the reciprocal of this.
For a full, infinitely defined β, this reciprocal law remains in a limit sense:
Ville’s Raw Criteria:
S is a nil set $\leftrightarrow$ There is a β with infinite gain, that is success on S.
For the $\rightarrow$ direction first observe that if S is a nil set then by the previous result we can have betting functions that have arbitrary big gains. The next trick is to use a weighed combining of these betting functions and thus also weighed bettings must finally enter.
For the $\leftarrow$ direction we can regard the β over the longer and longer beginnings. Then again by our previous result the possible gains here have as maximum the reciprocal of the minimal cover of the beginnings. Now if S were not a nil set then these minimal covers would always exceed a fix ε total length and so the gains were always under $\frac{1}{\varepsilon}$.

8. My Own Approach And A Short History

We had five alternatives for S being a nil set. The definition as having a B cover pool for all ε with $< B > \le \varepsilon$ is the most visual but actually the most complex because it involves a huge infinity of possible pools for every ε. A trivial simplification is to have a diminishing pool sequence that covers S. But still we have a sequence of pools. Solovay’s Raw Criteria is requiring only a single pool. Then we showed how the quota can tell the nil coverability but this stayed with cover sequence. And finally Ville’s Raw Criteria used betting functions.
The Law Of Stopping was introduced at the start and I even showed why it should be a legitimate law of randomness. By this vision, continuing in a narrow pool is a strangeness and so Solovay’s Raw Criteria says that equivalently being in a nil set is a strangeness too. Ville’s Raw Criteria is then the triple whammy confirming that being in a nil set is a strangeness.
In 1965 in high school I realized the Law Of Stopping. I also believed that continuing in a narrow pool is a universal strangeness. That is, not continuing in any narrow pool guarantees a sequence to be random.
Strangely, all this was the result of a wandering away from the Goldbach Conjecture.
This claims that every even number is the sum of two primes. But this means that every number is the average of two primes. In other words, every number has a prime under and above same distance away. So if we place 0-s for composites and 1-s for primes, then in this sequence, mirroring a beginning will always make two 1-s collide. The falsity would mean that there is a beginning that when mirrored doesn’t collide at any 1 value so the double long beginning is actually “anti symmetrical” in this sense. Obviously this is a very strong feature and randomly chosen sequences can only show such symmetries or anti symmetries as accidental features that stop after long enough beginnings. The simplest is a repeat of a beginning.
I was amazed by how deep this stop intuition in everybody. I remember running around and asking all my family members about their feeling and all claimed correctly that infinite many times such repeating beginning would be absurd.
I knew that if a beginning property has a chance above any small value in all the B^n length groups then this property should repeat infinitely unless it is stopped by earlier beginning.

This is actually a use of the realistic Murphy's Law but the beginnings are not independent trials, they influence each other. For example, the property of starting with a 0 has half chance but once a beginning is such it remains and if it isn't then remains such. So, though we have a positive chance, the feature doesn't necessarily comes about.
This line of excluding the dependence can lead someone away from the real problem as it did me for a while too. This deeper problem is the question whether the reversal would be true, that is whether not having a fix "ε tunnel" as I called it, through which the random sequences squeeze through, would stop them appear.
I realized that the random sequences don't need such ε tunnel. They can come through a zero tunnel too. To be more precise, only fast diminishing properties mean stopping while slow diminishing ones can reoccur infinitely. I also realized the exact definition of this fast and slow as having a total of finite or infinite. So $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ is fast diminishing while

$\frac{1}{2}, \frac{1}{2}, \quad \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \quad \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \quad \ldots$ and

$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$ too are slow diminishing.

If the beginnings have the first as chances they must stop reoccurring but if they have the second or third then they can reoccur infinitely.
These two kinds of diminishings were introduced in the start as being concentrated or well concentrated and these are much better spatial names.
The above three sequences are very educational for showing why the first imply stopping while the second and third don't. Now I show these and so repeat some earlier arguments:

If the chances are $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots, \frac{1}{2^n}, \ldots$ then the end total after the $\frac{1}{2^n}$ member is

$$\frac{1}{2^{n+1}} + \frac{1}{2^{n+2}} + \ldots = \frac{1}{2^{n+1}} \left(1 + \frac{1}{2} + \frac{1}{4} + \ldots \right) \le \frac{1}{2^{n+1}} 2 = \frac{1}{2^n}$$

This obviously approaches zero but this sum is also a bound on the chance that at least one of the $\frac{1}{2^{n+1}}, \frac{1}{2^{n+2}}, \ldots$ chanced members would come about. Indeed, this means an infinite "or" combination and a sum of chances gives the "or" for excluding events.
Like the six possibilities of a dice that add up to 1.
A non exclusion can only decrease the chance of an "or" to happen.
So the "or" must have maximum this sum as chance.
Now, the infinite "or" also means that we have an occurrence at all after the $\frac{1}{2^n}$ member.
Having infinite many occurrence of course implies all of these "or" occurrences and so the chance of this claim (having infinite many occurrence) can not be bigger than any of those "or"-s' chance. But those approach zero. So only one number is that is not bigger than all those, namely zero.
So the chance of infinite occurrence is zero.
Then a final step is to claim that this zero chance now means impossibility.

The second chance sequence

$\frac{1}{2}, \frac{1}{2}, \quad \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \quad \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \quad \ldots$

can be the chances of the following experiment sequence. We throw first single coins twice.
Then two coins four times, then three coins eight times, and so on. And we regard as success an all head outcome from our simultaneously thrown one or two or three and so on coins.

These group numbers are exactly the "combinatorically expectable" numbers. Two trials for single coin, four trials for two coins, and so on. They obviously do not mean guaranteed success in the group and only some kind of chance values could be attached to have a success in a group.
The first thought could be the half value but this is obviously false already at the start because to have at least one head from our first two single throws has $\frac{3}{4}$ chance. Indeed, the opposite of the success is having no head at all which has one case as tail and tail again and so has $\frac{1}{4}$ chance.
Then in the second group again with this trick the chance of success, that is having at least one double head is $1 - \frac{3}{4} \text{ x } \frac{3}{4} \text{ x } \frac{3}{4} \text{ x } \frac{3}{4}$ because a failure of the group is to have no double heads four times in a row and no double heads has three cases namely ht , th , tt from the four possible.
To calculate this fraction is not trivial arithmetics but if you care to do so you'll see that it will be again above the half value. I do not care to calculate it because I rather prove that in general too, in every group the chance must be above half. To show this we must prove that:

$$\underbrace{\frac{2^n - 1}{2^n} \text{ x } \frac{2^n - 1}{2^n} \text{ x } \ldots \text{ x } \frac{2^n - 1}{2^n}}_{2^n \text{ many}} < \frac{1}{2}$$

To do this we can increase the second third and so on members by adding 1 to both the numerators and denominators and then show the claim for that product.
But then the numerators and denominators all cancel, except the first numerator and last denominator remains as $\frac{2^n - 1}{2^n + 2^n - 1} = \frac{2^n - 1}{2^{n+1} - 1} < \frac{2^n}{2^{n+1}} = \frac{1}{2}$.

To see even more clearly what we did, I show it for the second group:

$$\frac{3}{4} \text{ x } \frac{3}{4} \text{ x } \frac{3}{4} \text{ x } \frac{3}{4} < \frac{3}{4} \text{ x } \frac{4}{5} \text{ x } \frac{5}{6} \text{ x } \frac{6}{7} = \frac{3}{7} < \frac{4}{8} = \frac{1}{2}$$

What all this means is that we have more than half chance in every group to have a full head.
So by the Realistic Murphy's Law we must have infinite many successful groups.
Which of course also means infinite many successes in the trial sequence without groups.
And so indeed we got infinite success in a diminishing chanced outcome sequence.
So we do not need an ε fix chance tunnel by which the Realistic Murphy's Law guarantees the infinite occurrence.
The third chance sequence $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$ can be again regarded as groups but now starting with the single $\frac{1}{2}$ then two then four and so on members.
Then looking at the complementing chance values we get:

$$\frac{1}{2}, \underbrace{\frac{2}{3}, \frac{3}{4}}, \underbrace{\frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{7}{8}}, \underbrace{\frac{8}{9}, \frac{9}{10}, \ldots, \frac{15}{16}}, \frac{16}{17}, \ldots$$

The product is always $\frac{1}{2}$ because only the first numerator and last denominator remains.

And so their "and"-s, that is all group members coming about are all exactly $\frac{1}{2}$ chanced too.

And so their opposite outcome that is not all members happening has also $\frac{1}{2}$ chance.

But this means that in the group at least one member fails to come about. And of course, these were the opposites, so it actually means that one of our original members comes about.
Since we have infinite many such groups with all half chances, by the Realistic Murphy's Law we must have infinite many happening and so having infinite occurrence in our original sequence.
But observe that here in this argument we made a big mistake. This "and" combination with multiplied chances is only true for independent outcomes. In our previous example it was true but for any chance sequence it is not always true. Most importantly, the beginnings are never independent trials so all this seems to be irrelevant for our goals. That's why I was very happy to find an example of a beginning property that is again slowly diminishing and yet will come about infinitely in all random sequences. The diminishing is not that difficult. The infinite occurrence also follows from a simple argument we used already. Then most amazingly, I didn't have to prove the slow diminishing, that is the infinite total. Indeed, my belief in the Law Of Stopping implied that the chances have to be slowly diminishing. As a topping on the cake, I also saw that this property not only fails to be a nil set collection for the obeying that is continuing sequences, but actually is a 1 set, that is its complement, the stopping sequences are a nil set.
To see this beginning property we should call the "state" of a beginning the number of the consecutive 1-s at its end. Of course if it ends with 0 then its state is 0 too.
The "rank" of a beginning should be the highest number of consecutive 1-s before the last such consecutive 1 segment by which we defined the state above.
A champion is a beginning that just reached a new highest 1 segment, that is in which the state is bigger than the rank.
Adding a 1 to a beginning, the state increases by 1 while the rank remains the same.
Adding a 0 to a beginning, the state becomes 0 and the rank becomes the bigger of the state or rank.
From these it's not hard to show that the relative frequency of the champions is diminishing.
Now to see that there are infinite many new champions, we merely have to imagine arbitrary long windows and realize that there will be full 1 segments.
Finally, a failure to such behavior would imply a longest 1 segment with infinite occurrence.
Say as one million long and then after a point all these would be followed by a predictable 0.
But this would imply a diminishing cover of all obeying sequences.
This diminishing cover or nil sets as strangeness to me at that time seemed a much worse vision than the narrow pools as strangenesses. In spite of that the first requires merely belonging while the second means continuing as "obeying". Somehow the whole finite versus infinite seemed to me a deeper point then. So the Law Of Stopping was conceived by me as:

Random sequences are those that stop in all narrow pools.

Unfortunately, now comes the shock:

Every sequence has a narrow pool in which it continues!
And thus of course there could be no random sequence at all.
To show such narrow pool for an s sequence is quite easy.
We simply have to collect the beginnings of s itself as a pool.
Obviously, s is continuing in this pool and the total chance is $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$.
Our instinctive feeling is that this catastrophe happened because we used the s as given but a random s can not be given effectively and so we can not collect its beginnings.
So then a solution should be that we require our pools to be given effectively.
A machine or some kind of finite formal system must collect those beginnings that are in our narrow pool. And then the new definition of randomness is this.

Law Of Stopping As Randomness Criteria:
An s sequence is random if for every machine or system that collects beginnings, if the collected pool has a finite chance total then s has only finite many such beginnings. Or in short form:
An s is random if every derivable, narrow pool stops in s.
So we have three finiteness in this heuristic definition.
Being derivable means to have a finite system that collects the beginnings.
Being narrow means finite chance total.
Stopping means to have only finite many such beginnings in s.
The false timely vision of sequences usually reverses the action and so we say not that the pool stops in s rather that s stops in the pool. The true meaning is of course the same.

I didn't come to this beautiful definition because the effectivity didn't come to me as a concept.
Instead I tried to restrict pools as being explicit in some formal system. This approach is still not used successfully by anyone.
The Effectivity approach or algorithmic randomness as it is called today was invented by Church but he used this for Von Mises' observational sequences.
Kolmogorov used effectivity, that is machines for a totally new direction, namely to define complexity or incompressibility of finite segments. Unfortunately, using these as beginnings can not directly lead to a definition of an infinite sequence being complex all the way.
It was Martin Löf who showed this failure and he started a totally different approach to collect all possible strangenesses as nil sets. Of course, the same contradiction that I showed above for any sequence continuing in its own pool of beginnings, will happen for nil sets even more drastically. Indeed, all sequences are members in a nil set because the sequence itself alone is obviously a nil set. So Martin Löf required the nil set to be effective by requiring the covering pool sequence to be effective and also effectively diminishing.
The complexity approach at the same time overcame the difficulty that Martin Löf showed and in fact it turned out that the new fixed complexity definition of randomness is identical with Martin Löf's effective nil sets as strangenesses.
Schnorr resurrected Ville's betting function to be effective and showed that it too leads to the same strangnesses. So he effectivized Ville's Raw Criteria.
Solovay effectivized what we called Solovay's Raw Criteria but he never claimed the Law Of Stopping in its perfect form for single pool, rather regarded a pool sequence with narrow total.
So in the end we have four definitions giving the same strangenesses and the simplified Solovay Criteria suggests the simplest randomness criteria formulated above. All this should suggest that we obtained a unique randomness concept.
And yet the reality is quite different. It turned out that effectivity could be injected in other ways giving a whole arsenal of different randomness concepts.
The mentioned historical delays toward the simplest forms suggest already some hidden blind spots. And in fact the simplest possible effectivization came latest by Kurtz.
I already introduced the basic idea of this at the start as the Law Of Occurrence.

9. Machine Injections

Already the effective collection or derivation of a single pool becomes ambiguous because two alternatives can be regarded. The earlier mentioned two possibilities of collections as generating the beginnings from all possible inputs or recognizing them from all possible inputs, is not this duality because these two are identical. But the duality is related to these two methods.
Both of these require alteration sequences that lead to terminations or halt states.
For a complex machine these alteration sequences that come to a halt may be arbitrary long and so we can not conclude that a particular text is not generable as output or not recognizable as input just because we waited long enough. An other more flexible method of being sure about our particular machine collection would be if we had an other machine that halts exactly for the complement texts, that is for those where our machine doesn't halt. Then we could run the two machines parallel and if the complementing machine halts for our chosen text then we don't have to wait for our machine any more. But this is also impossible for a complex machine.
The simple reason is that a complex machine always collects such texts that the complement set is not collectable by any machine at all. This surprise can be a bit better understood if we regard a machine collection as a listing of the texts. The point is that such list doesn't have to be in increasing order. We may have as first text a whole novel, and yet the single word "yes" could be collected as billionth in our list. This is so because our system of establishing what we collect does not simply relate to the length. A quite short text may come out as output or can be verified as a desired input after only a very long examination. Indeed, we use as examination method the alteration sequences and these can complicate short texts while simplify long ones.
Now the simple but not obvious fact is that the previous problem of the complement is related to this non length increasing of the lists. Namely, to be able to list some texts in increasing order at once means that the complement set is also listable. Indeed, all we have to do is watch when in our L list the first n long text appears, that is all $n-1$ long ones just have been finished. Then, in our complement list we can place all the finite many $n-1$ long ones that did not appear. And of course, this $\neg L$ complement list will be increasing too.
So, to list some beginnings increasingly is more than derivability, it is "dual" derivability or selectability. Indeed, for any given t text, going in our two lists simultaneously, we will definitely encounter t and thus select in finite time where t is, in L or in $\neg L$.
In spite of this sharp split of derivability, at first it might seem that this should be irrelevant for randomness because we collect eventually not beginnings rather covered or continuing sequences and the following two fundamental facts are true:
For any derivable B pool there is a dual derivable B^* pool, so that the sequences that are covered by or continue in B are also covered by or continue in B^*.
In addition, this B^* has a total chance not more than of B.
By the aboves, enough to show that for any L listing of beginnings we can create an L^* increasing listing so that the same sequences are covered or continue.
The method is amazingly simple! Whenever a shorter member pops up in L we replace it with longer versions. Namely, as long versions where we are in length. But we avoid all new members or all those versions that already appear. So if for example the first member in L is ten long and then as second member we have a three long beginning then we replace it with all possible ten long continuations. But if this second member happened to be a sub beginning of the first then we would make a continuation that is already there and so we should omit this version. Also, if later a new ten long comes up that already were created then we omit that too.
This L^* should be called the dualization of L and similarly the B^* content of L^* is the dualization of B.
The aboves suggest an amazing further simplification to avoid all those versions that are continuation of any already created ones and also omit new members in L that are continuations of created ones. This actually is L^{*0} that is, the B^{*0} minimal set of B^* in increasing order.
This is an excluding cover of B just like the earlier used B^0 minimal sub pool of B was but it is effective from an L listing of B unlike B^0. Indeed, in an L listing of B we can not tell whether a sub beginning may pop up later.

This L^{*0} should be called the pruning of L because it is very similar to the already used pruning of the full set of beginnings to get excluding pools.
Obviously the same sequences are covered by and continue in B^* as by and in B because we replaced all left out beginnings with all continuations of a higher length. This step also keeps the total chance and then the omission of repeats only decreases the total.
The further omissions in B^{*0} decrease the total even more and this set still covers the same sequences as B did because we only omitted continuations. But of course the continuations in B do not remain in B^{*0} which is drastically obvious because there are no continuations in B^{*0} at all, since it is excluding.
Now we need a meaning for the effectivity of pool sequences.
This is fairly obvious as an M machine that for every n input creates a list of B_n.
If there is such machine then the $B_1, B_2, ..$ pool sequence is effective. If beside this,
$B_1{}^{*0}, B_2{}^{*0}, \; . \; . \; .$ is diminishing, then $B_1, B_2, ..$ is an effective diminishing pool sequence.
And the sets covered by such pool sequences should be called effective nil sets.
This suggests that we arrived at the Solovay effectivization but amazingly, these effective nil sets are not enough to do the job.
Indeed, in our proof of the $\rightarrow$ direction of the Raw Criteria we regarded a subsequence from our diminishing pool sequence so that the sizes are under $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \; . \; . \; .$ and thus the total combined set of these is narrow.
In other words, we selected a fast diminishing subsequence from any diminishing one because only their total could provide us the single pool we claimed for continuing.
To obtain this pool we have to make the selection effectively and so we must know all the sum values effectively and know that they diminish. Or as a lesser option we must know some diminishing values that bound the actual sum values.
The first "know" means to have a machine that gives $< B_n{}^{*0} >$, while the second means a machine that gives a sequence of $\frac{N_n}{D_n}$ fractions with $< B_n{}^{*0} > \; \leq \; \frac{N_n}{D_n}$ and $\frac{N_n}{D_n} \rightarrow 0$.
The first is not possible in general and so we merely require such fraction sequence.
If we have such then automatically we can choose them to be in any concrete form like say :
$\frac{1}{n}$ or $\frac{1}{2^n}$, and in reverse any particular form implies all others.
Simply because we can select a sub sequence of the pool sequence.
We should call such pool sequences effectively diminishing and the sequence sets covered by them as effectively nil sets. So this "ly" ending tells that these are more than effective nil sets.
Assuming such effectively diminishing pool sequences, of course the $\leftarrow$ direction must also create such pool sequences. And unfortunately this is not true for our size group decrease construction because the size group chance totals are unpredictable. We don't have bounding fractions for these.
Luckily our other $B = B^0 \cup B^{(1)} \cup B^{(2)} \cup \; . \; . \; .$ perfect partition satisfied:
$< B^{(n)} > \;\; < \;\; \frac{< B >}{n+1}$.
And these are all excluding pools so their sizes are the same as of their prunings $B^{(n)\,*0}$.
We don't know $< B >$ but by this inequality the fix numerator $\frac{N}{n+1}$ fraction sequences will definitely be effective diminishing bound sequences for $< B^{(n)} >$ if N is above $< B >$.
Thus we can also choose any other concrete bounding fractions like $\frac{1}{2^n}$.
This was the original bound sequence chosen by Martin Löf for the effectively diminishing pool sequences and thus for the covered sets as effectively nil sets.
Though he didn't use this "ly" yet because he did not consider our effective versions.
Unfortunately, we overlooked that the effectivity of the members themselves is not granted at all.

Indeed, in an L list of beginnings we can only see that a beginning has a certain many sub beginnings if L is increasing. Otherwise new sub beginning can pop up arbitrary late in L.
So we have to obtain first the L* totalization of L and then use the perfect partition of B*.
Solovay used an other method for any L by combining the two earlier constructions that failed. Namely, using the perfect distribution but also the trick we used for the length groups, that is using a shrinking sequence by subtracting the members from the total. So:

$B_1 = B = B^0 \cup B^{(1)} \cup B^{(2)} \cup \ldots , B_2 = B - B^0 = B^{(1)} \cup B^{(2)} \cup \ldots ,$

$B_3 = B - B^0 - B^{(1)} = B^{(2)} \cup \ldots$ and so on.

Here we merely have to recognize if a beginning has at least certain many sub beginnings and this indeed recognizable by going far enough in any L listing of B.
But now of course the sizes of these shrinking members are much bigger. Luckily, they are covered by the perfect distribution members. So $< B_n{*}^0 > = < B^{(n)}{*}^0 > = < B^{(n)} >$.
In fact, the diminishing quota can replace the diminishing of $< B_n{*}^0 >$ too if we add all continuations to B_n to make the members complete. And this is effective too.
This cover sense was that Martin Löf used as definition in his effectively diminishing pool sequences and resulting effectively nil sets. Which is amazing, considering that he wasn't aware of the whole Solovay equivalence yet. For him this was also motivated in order to be able to use shrinking pool sequences.

The Martin Löf randomness should be regarded merely as a replacement of the Solovay.

The Law Of Stopping is the fundamental and absolute essence. The single pool is simply a bag of possible beginnings that has well concentrated chances, that is has a finite total chance.
The ordering of this single pool is irrelevant and the effectivity obeys this.

But the Martin Löf definition offers the mentioned alternative wider effective nil set concept that again should be used in cover sense. This wider concept of strangeness did become important and I will tell in the next section why.

It's weird that Solovay's definition came later, but even more weirdly the simplest possible concept came latest by Kurtz.
Here we have a single B pool that covers a measure 1 set of sequences and then any random sequence has to be inside, that is being covered by the pool. The left out sequences are a zero set as strangeness. The effectivization is easy by saying that B is listed or listed increasingly. And here again using total effectivity makes no difference as criteria of randomness.
Also observe that the required 1 measure of the covered sequences by B means that the B** pruning has 1 total. So we can replace the whole definition of Kurtz randomness as a reverse of the Law Of Occurrence in any totally effective excluding B with 1 total.

It's easy to show that the effective zero sets are all effectively nil sets too, that is the Kurtz strangenesses are Martin Löf and Solovay strangenesses too. So the Martin Löf and Solovay random sets are Kurtz random too. To see that some Martin Löf Solovay strangenesses are not Kurtz strangenesses is harder.
The following final twist in effectivity will explain why there are such.
As we mentioned, there is a universality for machines too and so an interesting question is whether the universal strangeness or expectability is universal in this sense too, that is can be obtained by a single machine. Martin Löf showed already in his original article that his definition is such, which of course makes the Solovay variation such too.
The Kurtz strangenesses or expectabilities are not universal in this sense.

10. Compressibility

The aboves suggest that there are alternative randomness concepts that ruin the heuristic simplified Solovay definition. But this is unavoidable because the Martin Löf version naturally brings in other versions. Historically this was not how things happened.
First of all, there was Ville's heuristic betting function that again turned out to be equivalent if we regard them to be effective, but this was forgotten and Solovay's variation came years later too.
Instead, the concept that turned out first to be equivalent with Martin Löf's effectively nil sets was an other Kolmogorov road and actually this initiated Martin Löf's definition.
Simply because Martin Löf was studying under Kolmogorov and he was the one who showed that this approach had a glitch. So he abandoned it and instead injected effectivity into Kolmogorov's first road that was the axiomatization of Probability leading to strong Laws Of Large Numbers that draw attention to nil sets of sequences. But the glitch became fixed very soon and the new second Kolmogorov road lead to the same randomness to what Martin Löf arrived from the first.
So what was this "second coming of Kolmogorov"? To be truthful the main idea was first introduced by Solomonoff but the world knew it from Kolmogorov. Probably he had this idea way back when he turned his back on randomness and axiomatized probability. And in a sense it is a drastic opposite of the whole probability vision of possible outcomes.
Up to now I tried to project the following vision:
Among finite possibilities the possibilities are all the same and the false feelings of chances being dependant on the past can be easily corrected by really calculating them. So a new head or tail is always merely halves the chances and thus after hundred heads a new head or tail is still equally possible. All fix long outcome sequences are equal. The all head or all tail or alternating head tails are purely subjectively special and any picked outcome sequence is just as "special".
Quite oppositely, for infinite outcome sequences certain strange ones are simply impossible! The strangenesses must stop!
The heuristic new vision of Solomonoff and Kolmogorov is the following:
Okay, the chances are the same in finite outcome sets but to claim that a billion long alternating head tails is the same as any "random" billion long outcome sequence is insane.
Namely, the simple alternation is not merely subjective as we claimed above!
This or any other property that "rules" a sequence can be used to describe it in a much shorter way than listing it. This is an effectivity problem and so machines can tell these subjective shorter definabilities or compressibilities quite objectively. Then what we should expect is that in spite of the ignorance of this by the finite chances, the earlier claimed infinite randomness somehow should emerge from this already finitely present compressibility or information content.
The name of this grand vision is AIT that is Algorithmic Information Theory and it was brewing exactly when I was in High School and got attracted to randomness. It is still a major mystery to me why I came to the simplified Solovay definition on my own and yet never even touched upon this alternative idea. In fact, I only absorbed it a decade ago.

This compressibility approach again uses machines but as output producers.
A machine can create from an input if it's used smartly as a program, a much longer output.
This long output then is not really random because the program describes it.
In reverse, if there is no such programmed compression of a long segment then it should be regarded as random. This was Kolmogorov's basic idea.
Of course, every machine can be enlarged or altered by adding irrelevant contents. So we could store in our machine some predetermined very long segments and so we could instantly produce these and so seemingly compress them as if they were dictated by a program. This wouldn't influence the infinite tendencies of compressibilities in a pool for a fix machine but using different ones we could allow bigger and bigger ones and so compress some arbitrary long segments. Luckily, as we mentioned, machines can all be simulated by a universal one and so choosing such, or even better choosing the minimal such, we get a perfect way of defining compressibilities in the long run.

This instantly offers the idea that infinite sequences should be regarded as their beginnings and some compressibilities for all of them should be interpreted as the sequence being strange.
Or in negative, a random sequences should have incompressible beginnings in the long run.
In the end, this was exactly what came out but not as easily as it sounds.
The glitch to sequences was realized by Martin Löf who then as an alternative discovered the effectively nil sets. Then at least three people discovered a way out of the glitch and succeeded in the incompressibility definition of randomness.
Amazingly, the “alternative” and the “way out” became the same, so the fixed up incompressibility definition of randomness turned out to be equivalent with Martin Löf’s.
The most famous of these glitch fixers became Chaitin who put his book on the net for beginners too and there he claims: “There is only one randomness, mine!”
And yet, since then the field exploded, and we have hundreds of alternate versions of randomness. In New Zeeland there is a whole group dedicated to this field.
Chaitin’s demagogue opinion is especially weird because exactly his ingenious major discovery lead most people to doubt the absoluteness of “his” randomness.
Chaitin realized that the finite < B > sum is an unbelievably smart number because from it we can recover much more than the machine itself that creates it.
Indeed, just because the machine creates this number it doesn’t mean that this number is predictable by the machine. Namely, this “creation” is merely a gradual reaching of a value. As our machine produces or recognizes more and more beginnings, we can calculate the chance and add it again and again. But the values of the partial sums tell nothing about where we are in a final sum or even if we will get to infinity.
The actual beginnings of this total sum are only knowable for a few digits at best.
The amazing consequence is that this number is actually a random sequence. And yet it is created by a single machine. Some would even think that it is then effective.
Of course, it isn’t by the exact definitions because they require not such approached determination by a machine rather a digit by digit determination even if not in the consecutive order.
Still, to be random is weird.
Most amazingly, the already mentioned widening of effectively nil sets to mere effective nil sets, will accept these sum values as strange and so with this as randomness criteria we don’t have such machine approached random sequences. But we are never out of trouble because we can always create other more complex machine determined sequences that are “random”.
Even more amazingly, the Kurtz approach, that is the Law Of Occurrence can also be generalized to give exactly these wider strangenesses in expectability form.
So, we can create very convincing coincidences for other randomness concepts too.

We might think that the whole problem lies somewhere in an imperfect injection of effectivity into randomness. I want to show that this is not the case! There is a problem with raw randomness already, without the effectivity injection and I already saw that in 1966.
I still believe in the same as I did then when I didn’t see the evils of social lies especially in the media and academia. This belief was that math is about plausibilities. That’s why I asked always naïve opinions about fundamental intuitions. I practiced what Hilbert said that if someone wants to understand something in his field he has to be able to explain it to the first man on the street.
As I said, my family members all agreed with what I found obvious too that exactly repeating beginnings can only happen finite many times. But as my personal road to the narrow pool showed, these plausibilities are actually false! Indeed, the stopping in narrow pools is totally irrelevant to the continuations of the beginnings. It happens purely as chance enforcement and happens for restarted longer and longer segments too. At the flights over the random chess board I explained this even more visually but I gave no justification for it. And I can not for sequences either. I did mention a logic, namely that if the chances dictate stopping that is for increasing segments too, then the beginnings can not alter this. But this has nothing to do with plausibilities. We do not have any plausibility about the fast or slow diminishings and yet we have very strong plausibilities that certain beginnings will stop.
A very interesting case was the mentioned “1 champion” beginnings.

Saying that a b is “1 champion” and also starts with 1 obviously halves the chances and so will remain slow diminishing. But it will only be able to continue if starts as 1.
So, half of the random sequences are stopping due to this barbaric cutting out.
An interesting task would be to find a slow diminishing property where the beginnings must always stop in it. It would be weird because non overlapping segments would have to continue in it by similar argument that I showed for the concrete reciprocal chances. So this property would stop in random sequences but not by the chances rather by the beginning influencing. Or negatively, its continuing would be a strangeness but not due to chances.
All this relates to the imperfect plausibility of our final definition of randomness.
Others recognized too that the conquest of Martin Löf randomness in universal expectability form as universal statistical test, is actually a mere empirical fact. So we simply have no counterexample by contesters just like at the Effectivity Thesis.