Continuing and Stopping

We finished the Sets book with the König Path Existence. It claimed that for any infinite B beginning set there is an s sequence that "sub-continues" in B. Meaning, that every s^n beginning of s is sub beginning in some b member of B. The stronger claim that an s exists that every beginning of s is actually a member of B is obviously not true.
In fact, if B is a " non-continuing" set that has no continuing pairs as members then an s can only have a single beginning in B. Anyway, this stronger claim that all beginnings of s are members in B, should be called as s step by step continuing in B.
But the sub-continue in B condition has an other formulation by only requiring that infinite many beginnings of s are sub beginnings in B beginnings. Indeed, this automatically implies it for all, because then every s^n beginning of s has a continuing s^N beginning in s that is sub beginning in a b in B and so s^n is sub beginning of b too.
This then offers a weaker but still strengthening of the sub-continuing by merely requiring that infinite many s beginnings are present in B. So we should call this as: "s continuing in B". Our previous remark about the non continuing B-s shows that an s existence is not always true for this weaker claim either.
If B is regarded as a beginning "feature" then this condition has an even better visualization. Namely, we should partition B into its different length groups as: $B = B^1 \cup B^2 \cup \; . \; .$
Then these B^n groups is what we really see as longer and longer cases of the feature that can show up in an s sequence. If s continues in B then of course we must see infinite many of the group cases being true in s and so we can also say that B continues in s.
The opposite is of course stopping and so then the s sequence will not have B kind of beginnings after a certain length. Being a B kind of beginning has a certain chance among a given length of beginnings. With increasing the length, the chance can vary and so if it decreases then we might feel that it stops occurring due to this diminishing.
A very simple opposite claim can be made easily too. Namely, if the chance stays above even a very small ε value then the beginnings as repeated trials force that it keeps reoccurring.
But this as criteria for the infinite reoccurrence has multiple problems!
First of all, the diminishing is not the opposite of such ε value existing.
Indeed, a diminishing is staying under any given ε from a point.
So the opposite is not having such staying value, rather popping up repeatedly for ever.
Then even if we assume that a chance diminishes, we can argue that the infinite many trials can accumulate rarer and rarer but still infinite many occurrences.
So we are close to some vital points but we must approach them more systematically.
At the same time our plausibilities will be tested as well. So we should start from here.
Namely, from an Elementary School level Paradox.

Exclusion Paradox

We regard the numbers up to ten million and exclude those that contain a single digit say 6.
How many numbers remain?
Most kids would think that the large majority remained while in truth less than half remained.
The proof is very simple. Up to ten we have one exclusion from ten and so .9 portion remained.
Then up to hundred we must exclude in all ten groups that were excluded up to ten so the 6 ending numbers but also a whole group, the sixties. So the remaining ones are $.9 \times .9 = .81$.
Then a whole hundred group will be excluded so up to thousand we have $.9 \times .9 \times .9 = .729$ portion remaining. As we see, the remaining portion keeps shrinking relentlessly in spite of the seemingly insignificant exclusion. And $(.9)^7 < .5$.
We feel that by increasing the value up to which we count the remaining numbers the portion will become less and less. And then we might say that regarding the infinite decimals then those that have no appearing 6 digit in them are merely a nil set. Strangely, a very correct statement but our argument had again more holes. First of all, how can a set be nil set but even more importantly, did we really prove that the ratio of the remaining numbers diminishes!

That is, multiplying $.9$ more and more times, that is $.9^n$ will approach zero?
Luckily this is quite easy, though not trivial! We can show that if $p < 1$ then $p^n \to 0$.
To see it, let $p = \frac{1}{1+\varepsilon}$. Then $\left(\frac{1}{1+\varepsilon}\right)^n$ will have in its denominator $1 + n\varepsilon + \ldots$
The more fundamental claim that the 6-less sequences are a nil set remains for later.

Segment exclusion and segment prediction

Instead, we turn to a more immediate consequence of our result concerning binary sequences.
Among these to exclude a digit is insane because then only the other digit can remain and so we end up having only a single sequence. But let's not give up yet because a salvaging of our result is easy if we regard a certain window or segment length like say ten digits.
Then excluding a certain combination say the alternating 0101010101 is again an exclusion.
Ten digits can have $2^{10} = 1024$ many variants and so excluding this single, is leaving now a portion that is $\frac{1023}{1024}$ so much closer to 1 than .9 was.
Still, the logic is the same and so excluding such groups will diminish the possible binary sequences to a mere nil set, though we didn't define them exactly!
In spite of that, it tells us that it should be something very rare.
And so randomly producing the digits by throwing a coin it shouldn't be possible.
Even this is now quite plausible because among the ten segments everything should come up.
But we pursue something else.
Namely, that by knowing that this 0101010101 can not occur for some crazy reason, would let us make sure predictions and so potentially make a fortune if someone takes our bets.
Indeed, if we know this "insider secret" that such combination of ten digits is impossible then waiting long enough, namely about 512 many nine long segments, we should get by chance the 010101010 segments too and then we could be absolutely sure that a 0 will follow.
So in these conditional situations we could make sure bets for 0. Slow but fix money!

Realistic Murphy's Law, An admission at the start

We had no doubts about our ten long window arguments.
In fact, we can even imagine that instead of throwing a coin ten times we could take ten coins in our fist and throw them simultaneously. The number of trials to get a desired outcome combination for our ten coins is again expectable to be 1024. A single throw of our coins and collecting them back into our fist takes a few seconds. So making throws continually for a few hours, will get this many trials and so will likely to achieve all possible combinations.
The surprise comes if we ask this same expectable time for twenty or thirty coins.
2^{20} is so big that this many trials could only be done in a month.
For thirty coins we need thousands of years.
The original paradox was raised by the inventor of the Chess who asked as reward from the King a single rice on the first square then double on the next and so on.
2^{64} is so huge that the rice would be more than the atoms in the universe.
This is very surprising to people and yet they all agree that having all 64 heads is also possible!
And then comes the grand twist by asking whether the outcomes in an infinite sequence could be all heads. Of course to ask it as throwing infinite many coins simultaneously from our fist, is adding an extra layer of problem. Indeed, we all agree that we can not hold infinite many coins.
So being forced to use the timely consecutive trials, should already tell to a philosophical mind that we are onto something very deep here. And indeed, there is no trivial a priori reply.
But before I dwell into this deep problem in a very personal way, I want to give a name to this correct intuition which says that everything that has some even if miniscule chance, will happen.
I call it the Realistic Murphy's Law.

Indeed, the original claimed that the buttered bread always fall on its buttered side but this realistic just says that if we drop our bread again and again then the mishap will happen.
This trivially implies that if we throw a coin infinite many times then both sides must happen.
In fact, we feel much more, namely that they must have some equalized infinity of outcomes.
This is the vaguest form of the Law Of Large Numbers and we come to its detail later.
But now I just want to stay with the extreme all head impossibility because it is like a thorn in a giant. It reveals a much deeper problem.
Just a few weeks after I returned to Hungary to retire, I was invited to my nephew's wedding where I was talking for hours with my niece who is a math teacher. She asked about my interest and so Randomness came in, about which I corresponded with my brother too. She knew that measure is used to define randomness but was totally unaware of the details and thus had her intuitions intact. So she said that by her gut feeling all possible outcome sequences are equally possible, including the all head one. For seconds I was confused. The logical reply that the Physical Law Of Large Numbers that claims a certain equalization in the infinite trial sequences or the even simpler step of using the Realistic Murphy's Law didn't jump in. Simply because I myself was earlier in this grand intuition. Nature doesn't distinguish our mathematical outcome sequences from any random one. How would the coin flips know how to behave to avoid these.
But I erased this intuition rather than keeping it alive. Like every good lie, it happened subconsciously. After the few seconds when I did realize how to make a false bridge between the intuitions, I of course convinced her that the tail having a half chance sooner or later must come about. So the second layer of lying stepped in too, behaving falsely. But in the next few days I tried to be honest at least to myself. Humor always grasps the dead serious stuffs best.
So here is a gem. The tutor explains some math subject to the student who says he doesn't get it. Then he tries again without success. Finally the third time he starts from the bottom and when he fails again he says: How the hell you don't get it when finally I understood it too.
The return to a naive denial of the Realistic Murphy's Law means that we reach down into our soul and say: Why couldn't happen that all coins land on head!
And then this will alter our whole expectations about a Randomness Theory.
Indeed, then we still can accept that an all head outcome is so rare that it will never show in our restricted experiments in a fix universe. Or we can even contemplate about this "never".
But more importantly and mathematically, all the strange outcomes that a simple Randomness Theory would regard as impossible, are similarly just extremely rare.
And then the distinctions among these are at once a new field. Using our eternal intuitions!
That is the positive side! The negative is that then any solution by a Randomness Theory that uses Effectivity to block out the strange sequences must be rejected at once.
And indeed, the all head outcome is just a particular effectively very easily describable outcome! Every other "random" sequence is similar by Nature or God. The coins do not know about our machines and behave not accordingly. So if the all heads is impossible then any outcome were impossible. This is the simple and irrefutable argument against the grandeur of Algorithmic Information Theory. And the opposite reality is clear and simple:
Just as the all heads is possible but very rare, the same way any individual sequence is very rare. This just means that if one happens then we must wait for a very long time in average to have it occur again. So expectable return is the only correct concept for any outcome feature.
And an outcome feature is simply a subset of possible outcomes with given length.
And so a particular outcome is just a simplest elemental feature.
This means that I deny everything that I will present from now on.
They are a mirage with new meanings later when such Random Set Theory is defined.

Borel's Monkey, Varying Window

Now forgetting these deep doubts about the Realistic Murphy's Law, we can show how to make it paradoxical in a much less exciting sense by simply using sufficient smoke and mirrors.
It was achieved by the Borel's Monkey paradox in the beginning of the twentieth century.
Borel regarded a typewriter that has much more possible outcomes than head or tail.
But again, all texts are just outcome sequences from trying the finite many possible keystrokes.

To create a random infinite text we tie a monkey to our typewriter and let him play.
We claim that the poor monkey will sooner or later type down the exact Bible as a segment.
And then of course this will happen again and again infinite many times.
The "resolution" of the paradox is simple, Let's count the exact n number of characters in the Bible and regard the consecutive n long windows in what the monkey types!
All letter combinations must occur, including the Bible.
Now we bring in a new aspect not by using texts as the smoke and mirrors, rather pushing the seemingly simple intuition of arbitrary long full heads, that is 1-s a bit further.
The used "arbitrary" long could mean not just choosing a fix length window and then seeing the consecutive windows creating all combinations infinitely but regarding longer and longer windows in a single sequence simultaneously. The windows overlapping seems like a problem but the infinite occurrence of all lengths with full 1-s solves this and we get that all lengths must occur in a single sequence. Amazingly, we can get this result indirectly too.
If there were a maximal n number of 1 repetitions in a sequence then there would have to be an $m \leq n$ maximal repetition that occurs infinitely. Then after the last $m+1$ many 1 repetitions we would have a "sure" 0 occurrence after every m repetitions.
So this were a nil set property by the exclusion paradox and so our sequence were not random.

Law Of Duality, Difference increase

The number of 0-s and 1-s are equalizing in the long run and as I already mentioned, this seemingly natural but rather foggy intuition is the Law Of Large Numbers.
Strangely, we also use this as a false throw away label for the Realistic Murphy's Law. We say that this or that in spite of being unlikely will come about by the Law Of Large Numbers.
Our previous result that arbitrary long full 1-s can come about is a bit attacking a naïve equalization but now we use a much more precise second intuition that helps to make an even deeper attack on the Law Of Large Numbers. This intuition could be called as the Law Of Duality because it says that regardless how the 0-s and 1-s grow in their numbers, it is impossible that one of them would remain always more than the other.
So their total must periodically become more and less than the other's.
This then also means that periodically, they must have exactly the same number.
And here comes in our window argument with a twist.
We imagine windows only after these perfect equalizations. The overlappings again become immaterial due to the infinite occurrences and so in these windows then again everything must come about including full 0-s or 1-s.
So the difference of 0 and 1 totals must also be arbitrary big.
And so a rough meaning of equalization is trivially not true. And indeed the Laws Of Large Numbers mean only equalization of the 0 and 1 ratios relative to the beginning length.

Champion Beginnings, Sub n-champion beginnings

The existence of increasing full 1 segments suggests that when a beginning has so many 1-s at its end that exceeds all previous consecutive 1-s then we should call this a 1 champion.
In fact, better to say just champion for short because I'll use the n-champion name for a champion that has n many 1-s.
For a beginning to be a champion is obviously pretty rare and yet we must have infinite many of them by the following heuristic argument: We have arbitrary n long consecutive 1-s.
Regarding the first occurrence for every n we at once get the increasing n-champions.
Even more interesting is an indirect approach! If the champions would stop with an n-champion then we had longer and longer beginnings containing maximum n long consecutive 1-s.
This feels rare too. And by having infinite many champions this feature must stop.
So we found two concrete features that are both rare but one continuous while the other stops.
So if continuing and stopping is indeed the point then two kind of diminishing must lie behind!

Finite Sum Paradox, Achilles Paradox, Infinite Sum Paradox

If we stand a meter from a wall and move a half meter forward we'll be a half meter away. Then moving a quarter forward we'll be a quarter away. So the initial 1 is actually half plus a quarter plus an eighth, and so on. To turn this argument into a proof:

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

Or to be even more precise:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^n} = (1 - \frac{1}{2}) + (\frac{1}{2} - \frac{1}{4}) + \ldots + (\frac{1}{2^{n-1}} - \frac{1}{2^n}) = 1 - \frac{1}{2^n}.$$

And so:
$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^n}$ approaches 1.

Which is abbreviated again as $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$

So infinite many numbers adding up to only a finite sum is hardly surprising but still could be called as the Finite Sum Paradox.
The Greeks used already smoke and mirrors and created a more involved paradox that involves time and a verbal trick. It is the Achilles Paradox:
Above, approaching the wall from a meter, we didn't ask about time. Obviously it depends on our speeds at every step. Using same speeds, the times are decreasing proportionally too.
So there is a finite total time after which we reach the wall.
At that moment, there were infinite many times when we were still away from the wall.

And indeed, if at noon the rain starts then arbitrary small times earlier there was no rain yet! There is no paradox in this.
But changing the "arbitrary small times earlier" into "arbitrary many times" is a tricky thing.
Because the second usually means forever.
Simply because usually the times are not decreasing.
Of course, a fix distance travelled in space easily shows that there the times must decrease too.
So to confuse us, the distance itself can be taken as changing too.
This happens if we are not reaching a wall rather somebody else going in front of us slower.
So there we have Achilles who obviously should reach the half as fast running opponent even if that has a hundred meter advantage at start.
But after hundred meter, that is at the opponent's start, Achilles will not reach him yet because that will be fifty meter away. Reaching this point that will be twenty five meters away again.
So we could say that the slower runner is "always" ahead!
Which of course is false because these times add into a finite total when Achilles passes him.
We abused the word "always". Clearing the verbal usage, the paradox is resolved.
It is still useful because it provokes the role of time.
By the way, $100 + 50 + 25 + 12.5 + \ldots = 200 \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots \right) = 200.$
So Achilles passes the opponent at 200 meters.
And if his speed is v m/sec then the passing happens at $\frac{200}{v}$ sec after the start.
But this paradox was actually a detour and missed a much more important, potential Infinite Sum Paradox. Indeed, the smaller and smaller distances or time intervals adding into a finite total may suggest that smaller and smaller values always add into a finite total.
But this is false! We can have decreasing values that add into infinite.

The simplest example could be made by starting with a fix number say 1 which added infinitely is obviously infinite: $1 + 1 + 1 + \ldots = \infty$.
Now let's cut these into more and more pieces, namely the n-th into n pieces:

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

So we had indeed decreasing numbers that add up to infinite.
An other version is using $\frac{1}{2}$ steps instead of 1 and cutting them into not n rather 2^n many pieces :

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

Not so interesting, but it leads to an easy proof of the very surprising fact that the sum of all reciprocals is also infinity:

$$\begin{array}{ccccccccccccccccccccc}
\frac{1}{2} & + & \frac{1}{3} & + & \frac{1}{4} & + & \frac{1}{5} & + & \frac{1}{6} & + & \frac{1}{7} & + & \frac{1}{8} & + & \frac{1}{9} & + & \frac{1}{10} & + & \ldots\ldots & & = \ ? > \infty \\
 & & \vee & & & & \vee & & \vee & & \vee & & & & \vee & & \vee & & & & \\
\frac{1}{2} & & \frac{1}{4} & & \frac{1}{4} & & \frac{1}{8} & & \frac{1}{8} & & \frac{1}{8} & & \frac{1}{8} & & \frac{1}{16} & & \frac{1}{16} & & \ldots\ldots & & = \infty
\end{array}$$

$$\frac{1}{2} \qquad \underbrace{\frac{1}{4} \quad \frac{1}{4}}_{\frac{1}{2}} \qquad \underbrace{\frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8}}_{\frac{1}{2}} \qquad \underbrace{\frac{1}{16} \quad \frac{1}{16} \quad \ldots}_{\frac{1}{2}}$$

So we found something crucial about how numbers can diminish.
They diminish fast if their sum is finite.
They diminish slow if their sum is infinite.

Zero-Infinity Law

An $a_1 + a_2 + \ldots$ infinite sum's total is actually the limit of its $a_1 + a_2 + \ldots + a_n$ beginnings section if that exists. So if the total sum is an F finite value then the beginning section approaches F. And thus with positive members, the $a_{n+1} + a_{n+2} + \ldots$ tail section total must diminish as we drop members from it.
The reverse is true too, that is if the tail is diminishing then the sum is finite.
Because, the "reverse" of the reverse is that if the sum is infinite then the tail can not diminish.
In fact, even more is true! Namely, that the tail must remain infinite as we drop members.
So a tail section in any positive membered sum is either infinite or diminishes.
This could be called as a Zero-Infinity Law.

Probabilities

The chance of some property is the number of all possible outcomes with the property divided by the number of all possible outcomes. If these outcomes are all equal chanced!
This last condition is crucial and ignoring it can be a source of mistakes.
Simplest example of such error is regarding throws of two coins and then saying that the possible outcomes are three as two heads, two tails or one head one tail. So then we would get one third chance for all these three possible scenarios. In truth, the one head one tail means two cases because physically we can have two versions depending on which coin was which.
Using two differently marked coins, this becomes evident. This highlights an other issue. Namely, why simultaneous tries are usually replaced by repeated trials. Indeed, throwing a coin twice we at once realize that an outcome of head then tail is different from tail then head.

So time brings in the trivial outcome distinctions even with actually using common objects.
In spite of this, we must realize that according to our present knowledge, time has no role in the chances! Even a whole infinite sequence of tries should be the same as trying them simultaneously. This then becomes a paradox!
In a sequence of tries we expect the heads and tails to become half half portions of the longer and longer beginnings. This is the simplest version of the physical law of large numbers.
In spatial version the same expectation is about the “beginnings” as left end sections of the simultaneous tries. So we can visualize beggars in an finite row and each flipping their coins.
Then it’s natural to contemplate that such equalization also means that we can rearrange the beggars so that the outcomes would become head tail alternatingly. This as infinite outcome would of course be impossible randomly, in spite of obeying the physical law of large numbers.
The correct implication is that this law is obviously insufficient to describe randomness but a fatal error in the whole vision is that the dependence on order is much stronger than we thought!
Even if we regard 0-s and 1-s that are now outcomes for our beggars throwing dices denoting the no six or the six outcomes, we would have an infinity of 0-s and 1-s. The 0-s were of course now much more frequent and only a sixth of the beginnings should be 1-s.
And yet to rearrange these into an exactly alternating 0 and 1 sequence is again possible simply by both being infinite many.
So the equalization had no role in our previous trickery and we can alter even the chances.
The bottom of these two paradoxes are still not resolved.
In a sense, changing the trials by knowing their outcomes is the same as changing the outcomes themselves. But while we regard this second as a trivial cheating, the first is accepted as a possibility or “could have been”.
This rearrangement paradox of random outcomes is relating to the infinity paradoxes of sets.
The simplest is that a sequence remains a sequence if we cut off a beginning.
A more tricky one realized by Galileo already is that the odd and even members are two halves but each a sequence on their own. So a half of an infinite set is same many.

With new attached meanings, these set paradoxes become deeper.
For example, even the simplest one element cut off can have a new twist.
Imagine beggars in a row again! Steal the coin of the first and say him: “Steal your neighbors coin and tell him to do the same”! Eventually, we created a coin from nothing.

Among finite many outcomes of course there is no problem about these at all.
We find it natural that all combinations should come about and also that the strangeness in a finite outcome is only our projection.
So as we mentioned before, a fistful of coins thrown down can land in any combinations.
And still! Walking in a casino, and seeing ten consecutive black spins on the roulette, we will somehow feel that the red should be now more likely.

More involved case counting errors

First let’s see an error again with the cases to calculate the chances of double sixes in throwing two dices, which was frequent at early gamblers.
The aimed property has one case $\{6, 6\}$ and we seem to have 21 cases by the following logic: We have five other double outcomes and we also have fifteen mixed outcomes as :
$\{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\},$
$\{3,4\}, \{3,5\}, \{3,6\}, \{4,5\}, \{4,6\}, \{5,6\}$.
Of course, we were wrong because these mixed outcomes constitute actually two cases each if we distinguish the two dice somehow like one being black the other red.
And we have to do this because the double cases are singular physically.
We would not have two double six for example with a black and red die.
So the actual number of cases are all possible ordered pairs:
$(1,1), (1,2), \ldots, (1,6), (2,1), (2,2), \ldots, (2,6), \ldots, (6,1), \ldots, (6,6)$.

The first members are the black die outcomes the second the red.

And of course we have 36 pairs. So the correct chance of a double six is $\frac{1}{36}$.

The really interesting situation starts if we stay with this set of possible outcomes but look for not a double six rather to throw at least one six.
The desired outcomes are: (6 , 1) , (6 , 2) , (6 , 3) , (6 , 4) , (6 , 5) , (1 , 6) , (2 , 6) , (3 , 6) , (4 , 6) , (5 , 6) , (6 , 6) eleven in number and so our chance is $\frac{11}{36}$.

Surprisingly big at first sight but then quite oppositely it seems not large enough.

Indeed, we might say that one die has $\frac{1}{6}$ chance and so at least one should be $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Or maybe even more because the double is "extra".
That something is wrong with this "logic" comes out if we apply the same for a coin.

Then two coins giving at least one head should have $\frac{1}{2} + \frac{1}{2} = 1$ chance which is obviously absurd. In fact, we know that we have three desired cases from the four.

Or in an even simpler way, we have one undesired case, the double tail. So our chance is $\frac{3}{4}$.

Accepting the $\frac{11}{36}$ chance for a double six, prepares us for an even stranger fact among the triple throws. Three throws is half of the six outcomes and so now a desired triple six is even more seems to be half chanced. But the calculation gives again a surprise.
The total number of cases is easy as all possible triplets 6 x 6 x 6 = 216.
The desired cases have three groups. All six has one case, two sixes has 5 + 5 + 5 = 15 cases as three possible non sixes, and finally we have 3 x 5 x 5 = 75 cases for the three possible single six because this means two five combinations.

Altogether 91 outcomes, giving the desired chance as $\frac{91}{216}$ which is clearly less than half.

Observe, that we used addition in the second case counting, while multiplication in the last.
The reason is that the three positions of the possible five outcomes were excluding cases while the five outcomes for the two positions in the last situation were simultaneous.
Also, the first was for an "or" combination while the second for an "and".
This goes into a similar logic of adding the chances for excluding "or" but multiplying for independent "and". For example the already used three even sides of a die are excluding and so their "or" is $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$.

But the double six is the "and" of the two independent outcomes and indeed $\frac{1}{6} \text{ x } \frac{1}{6} = \frac{1}{36}$.

Beside these two rules, we often use a trick of regarding not the desired, rather the undesired cases first as we mentioned above for the double tail.
For example, the previous 91 desired cases also can be obtained much simpler as the remaining cases of the undesired ones which are the triplets without any six and so their number is simply 5 x 5 x 5 = 125. And indeed, 91 = 216 – 125.
This level that we are now was about where the early probabilitists were.
They were actually gamblers who realized that simple common sense can be misleading and started to calculate chances by counting cases. But sometimes they fell back to false intuitions.
The most famous case was of Chevalier de Mere who knew the above three dice paradox and so knew that to get at least one six from three consecutive throws is also less than half chanced.
So if we play with double paybacks then betting on at least a six from three throws is a "looser".
But he also knew from playing a lot, that from four throws with same double pay we will win.
And indeed, there the possibilities are 6 x 6 x 6 x 6 = 1296 and the desired case number is:
1296 – 5 x 5 x 5 x 5 = 1296 – 625 = 671 > 625 and so the chance of a six containing four outcome is bigger than one without a six.

But then he used a logic to transfer these to sequences of double dice throws.
Here the six possibilities are replaced by 36 many and he thought that just as four consecutive trials overcame the half chance barrier, here 24 trials should do the same because
4 : 6 = 24 : 36. Unfortunately, he was wrong. The proper counting shows why!
The total cases of 24 throws are 36^{24} and the undesired ones are 35^{24} and so the desired ones are $36^{24} - 35^{24} < 35^{24}$. This can only be seen with a computer of course.
Amazingly, he realized that he was wrong by simply loosing and this was regarded as a contradiction in math. So he contacted Pascal for help who also discussed it with Fermat.

Physical Law Of Large Numbers

The fact that de Mere lost repeatedly proved that his chance calculations were wrong.
So the physical experiments tell the truth.
The simplest rule that we all feel natural is that the heads and tails even out in the long run.
But we never contemplate a fundamental contradiction in this expectation.
Namely, that experiments supposed to be repeated.
Even our experiences about nature are repeated observations.
Yet here at throwing coins again and again, we somehow trust this single outcome sequence.
This is due to the fact that a single such sequence is already infinite.
And indeed, if we for example regard the fix length windows in our sequence then those repeated outcome windows can be regarded as infinite many trials. So that's exactly why the Borel's Monkey paradox is completely resolved once we see that the Bible is also just a fixed window outcome. But this claimed equalization of the heads and tails is totally different.
It requires to regard longer and longer beginnings for the claim itself.
So indeed, the individual full outcome sequence must always behave the same way.

Weak Law Of Large Numbers

The early probabilitists just like the gamblers, wanted to put some math behind this whole thing. The possible infinite sequences were too difficult to investigate so they sticked to the beginnings. For example, you can make a list of all the possible ten long head tail beginnings.
There are exactly 1024 lines in this table and you can make a simple counting of those that you consider as important. But what should be the property you look for? Well obviously being close or far from an ideal half half equalization. So we introduce the e error of a b beginning as the difference of the 1-s portion in b from the ideal $\frac{1}{2}$ ratio. And we also set an ε threshold and collect the e values being at least this much.

If $<b>$ denotes the length of a beginning and $[b]$ denotes the number of 1-s then:
$e(b) = \left| \frac{[b]}{<b>} - \frac{1}{2} \right|$ and $E(\varepsilon) = e(b) \geq \varepsilon$ is the feature we should collect.
In general, for a B beginning set or feature of beginnings we use B^n to denote the n long members in B and so in our case we must regard the $(E(\varepsilon))^{10} = (e(b) \geq \varepsilon)^{10}$ set.
Of course the big question is what we should do with these ε erroring ten long beginnings.
Well, we should compare their number to the total number 1024.
So in general again, with regarding any B beginning set or feature, the proportion of the B beginnings among the n long ones is $\frac{\text{num } B^n}{2^n} = B/n$.
Where "num" means simply the number of elements.
Regarding our particular $E(\varepsilon)$ feature, this $E(\varepsilon)/n$ will be the at least ε erroring proportion among the n lengths. This is a simple counting among the possible combinations.

No experiments no trials. The best would be to use math and find a formula for this proportion depending on the chosen ε value. Then strangely, we are trying to make our task harder.
Indeed, we don't really care about ε because we believe that for any chosen ε value the proportion must tend to zero. In abstract form: $E(\varepsilon) / n \rightarrow 0$.
This was a grand step. Introducing a necessary complication just to show that actually the complication itself is immaterial.
More than hundred years had to pass when this simple claim of the fixed ε error proportion diminishing was finally proved. This reveals that obviously no simple formula existed that would calculate the proportion from ε. But as usual, the deeper layers remained hidden too.
Did this proof finally prove that the e error is itself diminishing in trial sequences?
Now with hindsight it is obvious that NO! An individual outcome sequence could be anything.
No proportions among the beginnings can prove what happens in a single outcome sequence.
But the situation is much worse! And with our hindsight of our earlier sections we see why.
Indeed, remember the champion beginnings. This C set is an actual counter example for the logic that $C/n \rightarrow 0$ would imply stopping of C in a random sequence.
To be more exact we must realize that $E(\varepsilon)$ is a continuous set of properties depending on ε.
If any of these with a certain well chosen ε_0 value would behave like our C counter example, then we had a more concrete contradiction. Indeed:
$E(\varepsilon_0)$ is diminishing just as C was and yet we saw that C continues in random sequences.
If $E(\varepsilon_0)$ would also do this then infinite many times would we see $E(\varepsilon_0) = e(b) \geq \varepsilon_0$ in a random sequence. Also, e diminishing means that for every ε there is a length $N(\varepsilon)$ that $<b> \; > N(\varepsilon) \rightarrow e(b) < \varepsilon$. In particular, $<b> \; > N(\varepsilon_0) \rightarrow e(b) < \varepsilon_0$.
So in the negative sense: $E(\varepsilon_0) = e(b) \geq \varepsilon_0 \rightarrow \; <b> \; \leq N(\varepsilon_0)$.
But there are only finite many b that $<b> \; \leq N(\varepsilon_0)$ and so also only finite many b could be where $E(\varepsilon_0)$ stands.

<u>Nil sets as special measure</u>

A sequence of intervals as point sets combined can be called a "cover".
It is a cover "of an S set" if all points of S are "covered", that is are elements.
Finally, the "covering" of a cover is simply the total of the lengths of the intervals.
This does not reflect at all what a real length size of a covered S set should be because the intervals can overlap. But if we strive to get a minimal covering then of course we are "closer".
Unfortunately, a minimal covering is not always possible but a so called "lower limit" of any V set of positive real values always does exist. This is actually the maximal number that all the numbers in V are larger or equal than it. If V is the possible coverings of an S set then this lower limit could be called the "covering limit of S" and it is denoted as S^*.
But this is still not a perfect measure of S. For example, it is possible that S is inside $[0,1]$ and $[0,1] - S$, the complement of S is so intermingled with S that both S and this complement can only be covered by 1 totaling intervals. So $S^* = ([0,1] - S)^* = 1$.
This is a very extreme but important case of S and its complement being non measurable.
Measurability itself is very easy to define with the smart concept of the "over-cover".
This is the set of points in a cover of S that do not belong to S.
In other words, these are the "wasted" points in a cover of S.
If for every ε we can make a cover of S where the over-cover is coverable by ε totaling intervals then S is measurable. So then the wastes can be arbitrary small.
Measurability is complementing and then the two covering limits are also complementing.
The arbitrary small coverability used in the definition itself can be used on its own as a special case of measurability. Indeed, if S is coverable by arbitrary small totaling intervals then it is measurable for sure because the over-cover is automatically coverable by arbitrary small total.
In this case we say that S is a nil set and write $S^* = 0$.
But observe that in the definition of measurability we can not replace the condition with the requirement that the over-cover should be a nil set!

Indeed, the over-covers are different sets and these have to be arbitrary small not a fix set.
Being a nil set can have more concrete conditions. This is the line that leads to our subject.
In spite of its simple definability, measurability is a complicated affair!
In fact, even the S^* covering limit is not so easy to prove for seemingly simple cases.
For an I interval for example, we would instantly say that I^* is obviously the length of I.
Indeed, I covers itself. But remember that I^* is the covering limit, so to prove that I^* is the length of I we must prove that I can not be "under covered" by more intervals.
Cutting an interval into pieces and rearranging them can show amazing surprises.
Namely, gaining or loosing many points. And after seeing these, we wouldn't be so sure.
But luckily the no under coverability can be proven and even a B set of intervals has a measurable total length that is easy to define.
The used B letter suggests that we will regard beginnings as intervals which is not obvious.
We'll come to a clearer justification soon, but first we'll make a little detour.

Infinite decimals

The sequences as infinite binaries can locate the points of $[0,1]$ just as the decimals.
This is the most important consequence of the "Arabic Number System" that of course was never Arabic rather Indian and the Arabs only brought it to Europe.
The point in the base ten "Arabic" numerals was already not the base being ten, rather that there was a base at all. So that the numbers are made from powers.
Interestingly, this requires the use of zero and so the Pope who wanted to forbid the use of the new number system since it is relying on the Devil, was unknowingly very correct.
Later, this system was easily continued to its real success to cover fractions too.
This is pretty well taught in Elementary School, so the kids in the modern age have something incredible in their hand that would have made poor Euclid cry from joy.
Do realize that the Greek mathematicians struggled with showing that not all pairs of distances are fractional in their proportion, or as we call them officially, "rational". This is an immediate consequence of the infinite decimal description of $[0,1]$ if we also observe something that is again taught pretty well in Elementary School, namely the division process in decimal form.
We bring down more and more digits from the number to be divided and once all digits are used we bring down zeroes:

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

110
120
80
100
20
60
40
120

As it is marked, we reached a re-occurring remainder, the 12.
So from that re-occurrence everything just repeats again and again.
Such re-occurrence always must happen because the remainder can only be maximum as many as the divider. And so a fraction is always a periodic infinite decimal. And so trivially we must have infinite decimals and thus points on the line too, that are not fractional.
Unfortunately, this grand consequence is not mentioned in Elementary School, neither the smart method to calculate back from any periodic decimal the fraction.
In fact, all this remains a black hole through out in High School and even in tertiary education.
There is a second black hole about the whole correspondence of decimals to points!
Namely, that it has a "glitch" that is related to the fractions as periodic decimals.

The simplest period is a single zero and then we don't even write out the infinite many zeroes. We just call it a finite decimal and never mention that the point it locates can be located by an other way too using all 9-s: 2.807500000 . . . = 2.807499999 . . .
Using infinite binaries, this glitch is even more striking, for both possible digits as endings.
This suggests something deeper beneath and it is the third black hole in the education about the infinite decimals. Namely, that the finite decimals are envisioned as the zeroes omitted and so locating points. Yet the glitch of the all 9-s would be perfectly resolved too if we mentioned an alternative representation of the finite decimals as the small sub intervals by repeated ten divisions. At binaries too, the beginnings should mean the closed halving intervals.
I mentioned all this in the last section of the Sets book and showed how the König Paths Existence then becomes very visual because s sub-continuing in B means that the intervals corresponding to the beginnings of B approach the P point corresponding to s.
I even mentioned that if B is a non continuing set of beginnings then it means that no intervals are inside each other but an approached P is still there, corresponding to an s that will not continue any beginning from B and also not being covered by any interval in B.
To regard closed halving intervals as beginnings was useful to avoid the glitch of the endings but it was also important so that when we do have nested intervals, we have a covered point.

Cantor Space

Unfortunately, there is a new glitch because by using closed halving intervals, the middle points are present in both halves. To fix this naively is easy.
We cut [0,1] at the middle and pull the two $\frac{1}{2}$ long intervals apart also $\frac{1}{2}$ length away.
This $\frac{1}{2}$ long middle interval is regarded as a blank zone and our set will not have points there.
The real problem is that the point $\frac{1}{2}$ can only go either left or right.
So we add a new point and thus make both halves have perfect end points.
In the two halves we halve again and now we make $\frac{1}{8}$ long blank zones between them.
Observe that then in total these are $\frac{1}{4}$. We also add new points to make ends there too.
Next we must cut four intervals and we use $\frac{1}{32}$ long blank zones totaling $\frac{1}{8}$.
The number of new blank zones are 1 , 2 , 4 , 8 , . . with lengths $\frac{1}{2}$, $\frac{1}{8}$, $\frac{1}{32}$, $\frac{1}{128}$, . .
And the new totals are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, totaling 1.
So our original unit interval became 2 long if we count the blank zones too.
But the whole point is that we don't want to regard that blank zone and use only the infinitely dissected original intervals as our "space" where the infinite sequences now perfectly continue.
This is called the Cantor Space and Borel realized first how this is a visualization of the binary sequences and their beginnings as chances. Kolmogorov used it for investigating his new probabilities that can be assigned to sets of outcome sequences.
A sequence obviously has only zero chance but a whole set of points can have positive chance.
But even more importantly, even a whole set of points can have zero chance.
Observe an interesting duality! In a sequence of outcomes the half chances multiply into the zero total because the outcomes are independent and we regard their "and".
But for a set of sequences the chance means the "or" because we regard one member becoming the outcome for a sequence of trials. For excluding events the chance of "or" is the simple sum and since all sequences are different, these are excluding each other.
But what should infinite many zero add up to? Should it depend on how big infinity we use?
Well, infinities were discovered by Set Theory but how big the infinity of all sequences is still remained a mystery by the Continuum Hypothesis.

Nil sets as a Set Theoretical concept

Surprisingly, this is very easy to define. Namely, for an arbitrary A set we can regard possible binary evaluations of A as possible coin flip outcomes at the elements of A.
Usually this set is denoted as $\{0 , 1\}^A$ because for an n membered finite set there are 2^n many possible binary evaluations. For example, ten coin flips have $2^{10} = 1024$ possible outcomes.
Regarding the S subsets of $\{0 , 1\}^A$ we should have some chances of these.
The full $\{0 , 1\}^A$ evaluation space has obviously 1 chance or certainty, because every individual outcome possibility on A is inside this.
If we chose an a member of A and regard the S set of outcomes that have all possibilities except on a we fix 0 or 1 value then the chance of S should be $\frac{1}{2}$.
This can be generalized by regarding any w finite subset of A and fix the outcomes there.
The w stands for window and we already used this notation in the Sets book at the Taboo Avoidance. We even introduced $< w >$ as the number of elements in w and mentioned that the possible evaluations of w is $2^{<w>}$. Now this means that the chance of the W set that fixates some values on w is $\frac{1}{2^{<w>}}$. The W set can be visualized as a bundle of individual outcome cases that is tied around by the evaluation of w. This bundle then can be identified with the evaluation on w and so an S set of evaluations that is inside W can be also said to be “covered” by W and so also by the evaluation of w. From here the continuation is easy!
We can regard any sequence of window evaluations and then their chance total is calculable.
Also an S set is covered by such sequence if their combined bundles contain S.
An S set is nil set if for arbitrary small ε there is a cover sequence with total under ε.
Finally, this is the same as having a sequence of cover sequences so that their totals diminishes.
Even an opposite approach can be established in the A arbitrary sets.
Namely, not combining bundles, rather taking them out from the full $\{0 , 1\}^A$ set.
Cantor did this to obtain his famous Cantor sets by cutting out a middle section of $[0 , 1]$ then again cutting out middles, and so on. The earlier mentioned Cantor Space comes from this too.
Unfortunately, in an arbitrary $\{0 , 1\}^A$ space we can not see that the remaining set of a cutting out has particular covers. If however $A = \omega$ is the set of natural numbers then the ordering gives a heuristic simplification by using the beginnings as standard windows.
A very logical generalization would be using well ordered sets and it is not yet worked out.

Nil sets among the binaries

For the naturals as base set of course we have the binaries as points of $[0 , 1]$.
But as I just mentioned the big simplification is the use of beginnings.
Any w window has an n biggest member and we can then regard instead of w the extensions of the members up to n. So the evaluations of w are possible n long b beginnings.
The length of the interval corresponding to the b beginning is $b^* = \frac{1}{2^{<b>}}$ where $< b >$ denotes again the “length” of b , meaning the number of digits in it.
This b^* is also the chance value of any $< b >$ long beginning as outcome.
Observe that though the different b beginnings mean different halving intervals, their lengths are the same for same length b-s and also, these b-s have the same chance values as outcomes.
The ΣB total of a B beginning set is simply the sum of the beginning chance values.
The B^* total length or chance value is smaller because as we mentioned the intervals overlap.
The chance is also smaller than the total because the beginnings as trials are not independent.
The beginning vision is very useful to find B^* without the overlappings or chances.
A b beginning in B is minimal if it has no sub beginning in B.
The set of these is the B^0 subset of B and the total length of this, that is ΣB^0 is B^*.

By the earliers, a simple criteria for being a nil set can be formulated as follows:
$S^* = 0$ if and only if there is a sequence of $B_1, B_2, \ldots$ beginning sets that:
1. $B_n^* \to 0$.
2. Every B_n covers S, that is:
   Every s sequence in S continues some beginning in every B_n.

We get a valid criteria too if we replace 1. with $\Sigma B_n \to 0$ because $B_n^* \leq \Sigma B_n$.

Now we give an alternative criteria using a single B beginning set:
$S^* = 0$ if and only if there is a B beginning set that:
1. ΣB is finite.
2. Every s in S continues in B.

For the if direction:
We'll use the $B = B^1 \cup B^2 \cup \ldots$ length group partition of B.
$\Sigma B = \Sigma B^1 + \Sigma B^2 + \ldots = v$ means that $(\Sigma B^1 + \Sigma B^2 + \ldots + \Sigma B^n) \to v$
and so $(\Sigma B^{n+1} + \Sigma B^{n+2} + \ldots) \to 0$.
Also, 2. means that every s in S has infinite many beginning in B which implies that:
every s in S continues some b in $B^{n+1} \cup B^{n+2} \cup \ldots$
Finally, observe that: $(B^{n+1} \cup B^{n+2} \cup \ldots)^* \leq \Sigma B^{n+1} + \Sigma B^{n+2} + \ldots$
So the tail sections are a diminishing cover sequence for S.
For the only if direction:
We must find such B from the $B_1, B_2, \ldots$ diminishing covering sequence.
First the good news:
The diminishing implies that the sequences covered, will continue in $\bigcup(B_n) = B_1 \cup B_2 \cup \ldots$
To see this, we should make two useful definitions:
minB = the minimal length with which there is b in B.
$B_1, B_2, \ldots$ is longening if $\min B_n$ becomes arbitrary big as n grows.
And now an easy claim:
If $B_1, B_2, \ldots$ is longening then any s that is covered by $B_1, B_2, \ldots$ will continue in $\bigcup(B_n) = B_1 \cup B_2 \cup \ldots$ Indeed, longer and longer beginnings of s must be inside.
Finally, observe that diminishing implies longening and so $B_1, B_2, \ldots$ is longening.
Now the bad news: $\Sigma \bigcup(B_n) = \Sigma B_1 + \Sigma B_2 + \ldots$ is not necessarily finite!
For example as: $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \infty$.
So we'll make an other B'_n beginning sequence from B_n as follows
Let B'_1 be the first member in $B_1, B_2, \ldots$ that $\Sigma B_n \leq \frac{1}{2}$.
Then B'_2 be the first that $\Sigma B_n \leq \frac{1}{4}$. And so on.
This new B'_n is again diminishing and covers the same sequences. Thus by our first good news, the sequences covered originally will continue in $\bigcup(B'_n) = B'_1 \cup B'_2 \cup \ldots$
And now the really good news: $\Sigma \bigcup(B'_n) = \Sigma B'_1 + \Sigma B'_2 + \ldots \leq \frac{1}{2} + \frac{1}{4} + \ldots = 1 =$ finite.
So our wanted B is $\bigcup(B'_n)$.
This new criteria could also be called as the Raw Solovay criteria because later Solovay injected effectivity into it and thus grasped the simplest Randomness definition.
It is strange that it wasn't spelled out earlier for nil sets.
It is also important because it makes a relation between fast and slow diminishing.
Indeed, S being a nil set just means that a diminishing cover exists for S.
But in our corresponding B the ΣB total being finite also means that the ΣB^n group chances diminish fast. So those s sequences that are covered by diminishing B_n interval sequences are also the ones in which some B feature's ΣB^n group chances diminish fast.

In fact, observe that these ΣB^n group chances have an earlier used meaning too as the B feature's B/n proportion ratio. Indeed:

$$\Sigma B^n = \sum_{b \in B^n} b* = \sum_{b \in B^n} \frac{1}{2^n} = \frac{\text{num } B^n}{2^n} = B/n$$

I think that the fast diminishing of these B/n proportions in an s sequence and yet B still continuing in s is the hidden intuitive trigger to feel that these s sequences are very rare.
In High School when I first started to see behind randomness, I was amazed how this hidden judgment of the fast diminishing is present in everybody.
I tested my relatives for different B beginning features and asked if the feature would have to stop or continue in a random sequence. They all got it right, all the time!

Strong Law Of Large Numbers

Kolmogorov proved 1. for the $B = E(\varepsilon)$ beginning sets.
In other words, that $E(\varepsilon) / n$ diminishes fast. And thus, and most importantly also, that:
Those s sequences where the ε erroring beginnings continue are merely a nil set.
This well sounding wording of his result, especially with the stressed "merely", should make us question whether we actually know more by this vision of using nil sets.
We know one big minus for sure! Namely, that no matter how we say anything about the sets of sequences, they can not deduce anything about the coin toss or random outcome sequences.
Simply because these are not defined at all. So the Physical Law Of Large Numbers is still out of reach. It could only relate to some Randomness Theory.
Actually, Kolmogorov was watching very closely the turbulent randomness defining efforts in the decades before the war. The main character was Von Mises who tried to use exactly the Physical Law Of Large Numbers. We'll come back to him but the point is that Kolmogorov realized that in a sense the chasing of Randomness was a wasted distraction from the "point".
And the "point" he found, was the nil sets.
But we should ask again! Did this nil set formulation bring in any new consequences?
The answer is a definite YES and I show now one simple such consequence.
It follows from the general fact that a sequence of nil sets combined is still a nil set.
The argument is amazingly simple, almost goes back to the early Cantor discoveries.
$\varepsilon = \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + ..$ So if S_1, S_2 .. is the sequence of nil sets then we should cover them by $\frac{\varepsilon}{2}, \frac{\varepsilon}{4}, ..$ totaling intervals which then together covers the combined set and has only ε total.
Now comes the second step, applying this to our situation where the sequences that keep erroring at least ε are always a nil set. We can imagine now any diminishing sequence of ε values and they don't have to be the previous halving values at all! The point is merely that we create a sequence of diminishingly "erroring" sets of sequences. By simply calling as "erroring" the continually erroring ones. And now comes the "point" that:
Their combined set will contain any sequence that errors repeatedly with any fix ε value.
Indeed, some ε_n member in our sequence must be smaller than ε and so the infinite many beginnings that surpass ε in their errorings will surpass ε_n too.
So the set of sequences that error repeatedly with any fix ε value are still just a nil set.

Zero-one Law

This claims that for certain S sets of sequences S^* can only be 0 or 1.
Namely, S must be such that for all s sequences in S, replacing any b beginning of s with any other b', the new s' sequence is again in S already. This should be called as S being "beginning independent" but the usual name is being a "tail set".
To get more realistic meanings behind, we could regard S set of sequences defined by some claims about the s sequences that are in S. Then this could be called a property or feature of s but the word feature we already used for b beginnings so we use here the word property.
Observe that while for the feature meaning we had the length groups as an abstract bridge, here among the infinite sequence we have no such. And this lies behind why properties of sequences are so hard to visualize as S sets. But let's get back to the beginning independence!
The $P(s)$ property is beginning independent if it remains true even if we replace any b beginning of s with any other b'. This sounds a bit more concrete and you can think many examples like having finite or infinite many 0-s or 1-s in s.
But most importantly, whatever we may come up as a randomness definition, such beginning independence is a "must"! Indeed, in a random sequence the beginning can be anything.
Observe that in our definition of the beginning independence we used the word "replace" and not "alter". Indeed, "alter" would suggest that b' is the same length as b.
Such "alterable" sequence sets are very important too but they are not enough for this claim.
We already mentioned that the S^* covering limit is not necessarily a correct measure and even mentioned the possibility that both S and its complement can have 1 covering limits.
The zero-one law does not say directly anything about this but its consequences do.
We'll come back to that after we prove the law itself later.
We also mentioned that measurability is complementing and it is not a trivial fact.
Neither is that then the measure itself is complementing too.
Beginning independence is complementing too which is only trivial by sheer logic.
Indeed, if S already has all the beginning variants of an s member then the complementing $\neg S$ set has to be similar. Indeed, for any t member in $\neg S$ if a t' beginning variant wouldn't be in $\neg S$ then it were in S but then since t is also a variant of t' it were there too.

The aboves show that the zero-one law allows three scenarios for a beginning independent S:
$S^* = 0$. Then $\neg S$ is automatically measurable and $(\neg S)^* = 1$.
$S^* = 1$ and S is measurable. Then $(\neg S)^* = 0$.
$S^* = 1$ and S is not measurable. Then $(\neg S)^* = 1$ too because $\neg S$ is beginning independent too and so it can only have 0 or 1 covering limit but 0 is impossible because then S were measurable.

Strangeness versus Expectability

Above I mentioned that randomness is very plausibly beginning independent.
Going deeper into informal naïve concepts, we can "define" that a strangeness is something that should not be true for any random sequence so is something that refutes randomness.
An expectability quite contrary, should be something that is true for all random sequences.
Which one should we use to define randomness?
Should we collect so many strangenesses that those sequences that avoid all of them are automatically random already, or should we collect so many expectabilities that those sequences that obey all of them are automatically random already.
Now observe that if a $P(s)$ sequence property is a strangeness then $\neg P(s)$ is obviously an expectability. So what the hell are we talking about as two different approaches?
The meaningfulness of this choice comes out if we go a step deeper into what we claimed that randomness is beginning independent. Every expectability must be such too because an expectability as sequence set contains all random sequences.
But of course a strangeness doesn't have to be beginning independent at all.

So to look for expectabilities we do have a guideline while for strangenesses we don't.
In spite of this, we have some a priori intuitions about strangenesses that completely disregard this most basic guideline.
If for example we see that the outcomes alternate as 0 , 1 , 0 , 1 , . . . then we instantly see a rule and know that this couldn't happen by chance, But this is not a beginning independent property, in fact changing the first digit to 1 the perfect ruling is destroyed.
Only further thinking will clear that actually an alternation from any point is impossible.
But then we are puzzled that any concrete "point" fixed, still leaves the property non beginning independent and so the "from any point" can only be collected as the infinite many possible starts of the alternation. So playing with these ideas more, we feel that actually strangenesses are deeper than expectabilities.

<u>Nil sets as strangenesses and Turing</u>

Exclusions are intuitive strangenesses and yet their nil set size is not plausible at all.
As our Exclusion Paradox showed it at the start.
So we must rise above intuition and accept abstraction as our guide as it happened in all modern science. Then we come to that "somehow" nil sets should be the strangeness.
That this "somehow" is a big problem, becomes evident if we just try to say that being in a nil set in itself is strangeness. Indeed, observe that any single sequence on its own is a nil set.
So then all sequence would become strange and no random remaining. So we must tell what kind of nil sets are to be regarded as strangenesses. That's where effectivity stepped in, actually the second time and so the idea of effective nil sets as strangenesses was conceived.
The combined adjectives, effective and nil is very misleading here. It suggests that just as the nil adjective that refers to sequences, the effective refers to sequences too. And in a sense this is true so we could talk about a sequence being effective but this is ambiguous and this ambiguity lies at the heart of effectivity. The ambiguity itself is very concrete: The 0-s or 1-s in a binary sequence can be effective on their own and one being effective does not mean that the other is effective too. This sounds strange first, after all if we know effectively the 0-s then we should know the 1-s too. But this is not true because knowing the 0-s effectively actually means to know their positions. Now if we can list the positions of the 0-s increasingly then obviously we can list the positions of the 1-s increasingly too because for any position waiting till a 0 with bigger position is listed, we can be sure that this position must be a 1. But if the 0 positions can not be listed increasingly and so a small position can come only after much bigger ones then this strategy is not working! For a given position we can not be sure if it is a 1 even after listing arbitrary many 0 positions.
In fact, as it turned out, this asymmetrical effectivity of the 0-s and 1-s is the typical and this lies behind the crucial incompleteness results of Gödel, that initiated Effectivity at all.
If we regard the positions as what they really are, subsets of the naturals then the aboves mean that an effective set's complement is not necessarily effective.
But this crucial and fundamental vision can be ignored by a second link with Randomness.
This link is that instead of sets of natural numbers we regard sets of sequence beginnings!
From the Random side this is very logical. The independence of the randomness of a sequence from any beginning in it or added to it, is the most intuitive certainty.
But as an amazing coincidence, the best effectivity approach of Turing also started with these.
A binary beginning is of course simply a finite set of naturals namely the zeroes or ones in it.
And here these two sets are equivalent because a finite set is automatically effective.
But this was not the important vision rather a more general one that allows beginnings of not necessarily binary sequences.
So we allow a fix alphabet that we abbreviate as 0-z meaning "0 to z" , that may include all the digits and special symbols of a keyboard. The rule table that will define our effectivity will have m many lines if our alphabet 0-z has m many symbols. The number of columns is free so can be arbitrary large n and we use simple natural heading for these. The members in such table with m many rows and n many columns will be so called action triplets.

A triplet starts with an alphabet symbol from our m possible ones. Then comes a horizontal arrow → or ←. Finally a natural column number, so any natural under n.
A simple case of alphabet is the 0-9 ten digits and a simple already existing effectivity system is the digital calculations we learn in Elementary Schools.
Can it be that by being general enough we can obtain everything that is effectively calculable?
The answer is yes and the amazing fact is that this generalness needs very simple possibilities.
Namely, our triplet tables can define every effective calculation.
Only one ingenious new idea is missing. Namely, when to stop our calculation.
Turing called this as halt and so in one of our triplets the third member is not a column number rather the letter h. Here is a concrete table for the 0-1 alphabet:

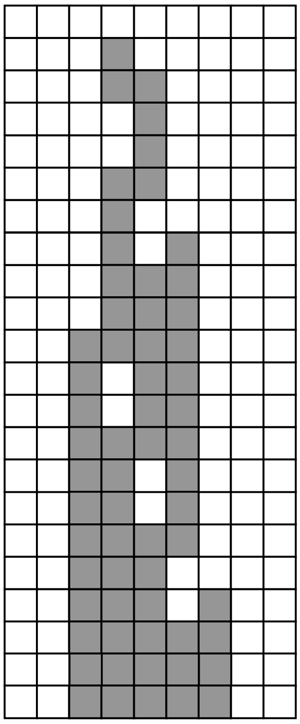
1 → 2	1 ← 2	1 ← 3
1 → H	0 → 3	1 ← 1

This table will tell an action sequence carried out on a squared paper.
Namely, we will write one 0 or 1 digits in every new line under a fully filled starting line.
Now we regard the simplest scenario when our starting line is all 0-s.
So we can start from any square of this number line and use our table to write a new number in the next line. Quite logically, we use the first column of our table to start our actions.
Remember, we said that the m many symbol values will also denote the lines in every column.
So our table has the 0 , 1 number values as line numbers in our first column too.
Not surprisingly, they will correspond to the value in the square, under which we write.
Or to put it an other way, this is the value we see above where we write.
Since we started from an all 0 line, we must use our 0 numbered, that is first line: 1 → 2.
The first of the triplet, 1 tells what to write in the new square.
The second, the → arrow tells that we should move one step in this direction and under this will be our new square to write in.
The third, the number 2 tells that we should now move in our table into the second column.
From here actually we repeat what we did previously.
We look up in our squared paper to see what the number value is above and go into this line in our table column. Since we started with all 0-s thus we see again a 0 above and so we go in our second column into the first line: 1 ← 2. So we again write a 1 in the new square.
But now we move left and so we will be under our previously written 1.
Of course, we must go a line down too and use the square under as next one.
The 2 tells that we must stay in our second column. To see what line to apply, we must look again up in our squared paper and we'll see a 1 so we must use the second line: 0 → 3.
Thus we write a 0 , move to the right and under this will be our new square to write in.
The 3 tells that we should move into the third column and we'll use the second row since we see 1 above. That triplet is 1 ← 1 and so we write 1 again under the already seen 1 above. We move left and move in our table to the first column.
The squared paper by now will look like this:

```
. . . . 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 . . . .
                          1
                            1
                          0
                            1
```

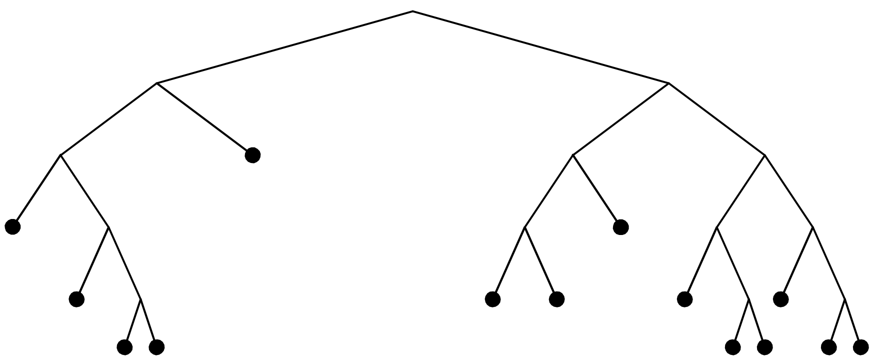
This could go on forever but observe that in our first column we have a H in the second row.
But do we necessarily encounter this halt order?
NO! And this is the whole essence of Turing's ingenious system.
Some tables from some starting line do get to halt while others from other start don't.
In fact, this double dependence of the halt from both the table and the starting line is an other crucial feature. It reveals that some starting lines can replace the tables themselves as programs.
This under-writing action sequence is very clear. We see the whole history, what happened.

There is only one negative feature, namely that the blank squares made this possible.
We saw through these blank squares to see the last written values above.
If we want to use the empty blank content as symbol or value too and for example we chose 0 as this then we must update every new line with all the filled in 0-s or 1-s.
Then everything becomes messier, so we'll show now all the steps till halt, with white meaning 0 and grey meaning 1. Even with this, to see where we were at a step is a bit confusing.
A simple help is that if two consecutive lines differ at a place then the writing happened there.
Our last halting place will be not such and so to help a bit we mark it with an x:

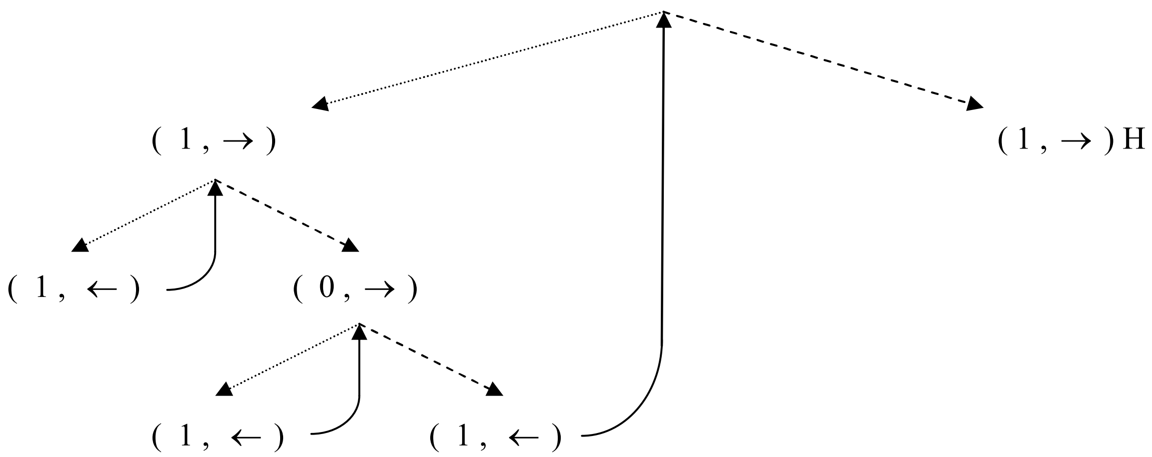


x

This updated blankless underwriting sequence was only pursued to come to the rewriting system that is the well accepted third step. Here we don't write the new symbols under the old rather we erase the old and write the new in its place. So we do not even need a squared paper just a line of squares. The elementary school solution could only be to use an eraser continually but the modern technical solution is to use a single line of memory cells!
This line being infinite means that actually we have an infinite memory and yet the single line means that we do not need memory registers and addresses.
The memory already suggests the further technical steps to turn a table into a machine.
We need a table head that does the jumps in the table and a memory line head that does the rewritings and single steps left or right. We also need a communication between them so the table head can get the last read symbol and the line head can get the new symbol and move direction. The almost unbelievable fact is that this system can imitate all real computers even if we attach to them an infinite memory. Now we'll show an "action tree" meaning for the Turing tables. First here is a finite graph with end points marked with dots.



Such graph becomes a redirected graph if we redirect the dots, except one.
This suggests already that this will be the halt while the redirected ones the other columns.
But this is not quite so because the merely continuing branchings are also columns that simply go to the next column.
Most importantly though, the two branching lines correspond to the read 0 or 1.
And finally, we must place the memory line actions to the branchings to get an actual action tree.
Here is the action tree version of the above table with solid lines as read 0 and dashed as 1.



As you see, we marked no states at all and actually in this case they are unique anyway.
Indeed, the top is column 1 and this determines the next as 2 which again determines 3.

If we would want to introduce these as special flow charts then it is amazingly simple too.
Remember that in general we have rectangles containing the actions and rhombuses containing the yes or no decisions. Here we make decisions prior any action but they are not yes or no decisions, rather for any possible number value. So we don't need rhombuses at all.
The rectangles are then the writings and movements, here in the parenthesis.

If it is a miracle that Turing's machines can perform anything that a machine can, then it is equally miraculous that the beginning collections by being recognized, that is halted by a Turing machine can define all those sequences that have anything strange in them.
But this is not the end of the story yet because we didn't tell exactly how the collected beginnings should define a nil set. Indeed, only a sequence of narrowing beginning sets defines a nil set. Of course one idea jumps in mind by the last condition for nil sets that we used to introduce the Strong Law Of Large Numbers.
It said that a single B beginning set can define a nil set if B has a finite total and by defining we mean those sequences that continue in B. This has a nice philosophical depth to it too because we actually used two "finiteness" about B.
The qualitative one is that it is machine collected and the quantitative is that it is narrow because the total chance values of the beginnings in it is only a finite value.
This Solovay Randomness definition is the most elegant and simple yet.

Von Mises

He was the key figure in trying to define randomness and his obsession was to widen the Physical Law Of Large Numbers, so that then it becomes mathematical to define randomness.
The fundamental problem is evident by the alternating $0 , 1 , 0 , 1 , \ . \ . \ .$ sequence.
This obviously obeys the Physical Law Of Large Numbers, in fact, too perfectly.
Von Mises said that a random sequence should obey it not only in its whole but inside too.
Namely, if we make observational sub sequences.
Now here in our alternating example, the every second members as observational sub sequence is full 1-s and so the Law Of Large Numbers is not true trivially.
The problem is that this concept of the observational sub sequence is indefinable.
But first observe that we shouldn't say observational "places" because if our goal is to exclude all strangeness then we must allow conditional place selections by earlier outcomes.
Indeed, after every three 0 having a sure 1 is an obvious strangeness and yet we do not know where these will be. We might think that simply the beginnings as condition of the place selection is enough because the past can not alter the next outcome. But this is still the same.
If I have a random sequence and regard all beginnings that are prior to the 1-s then I perfectly defined a B beginning set and yet the next outcomes are all 1-s.
So to say that after any given beginnings the next outcomes as observational subset must obey the Law Of Large Numbers is simply not true. Of course, this B set was not really "given".
That's where Church stepped into the picture. He was working on the concept of Effectivity.
And voila here a perfect application opened.
So he said that the "given" should mean collected effectively!
This was a turning point in the history of Randomness but the death blow came from behind.
De Ville realized that this ingenious help from Church is not enough.
The obsession with the Law Of Large Numbers as a fundamental feature is faulty in itself.
He proved that with this definition of Randomness, there will be sequences that though they obey the Law Of Large Numbers in all their effective observational subsets, they will fail the earlier mentioned other expectability, the Law Of Duality. In fact, there will be "random" sequences where in all beginnings the 0-s are always more than the 1-s or vice versa.

The six basic definitions of Randomness

If a B feature has a finite chance total and is collectable by a machine then the sequences continuing in B , are called "Solovay strange".
Without such effectivity condition, the definition would be making every sequence strange.
Indeed, the set of beginnings for any sequence has $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+ \ . \ . \ . = 1$ chance total and the sequence obviously continues in it.
Of course, these beginnings were totally hypothetical, merely the beginning of "any sequence".
The condition of machine collectable beginnings actually avoids this abstract sequence.
Strangely, accepting the machine collection we might feel that this in itself should be enough.
After all, a machine could not confirm a random sequence only a strange. But we are wrong!
Simply because we are not talking about "confirming" vaguely, rather "continuing in".
The best example to see our error is the set of all beginnings that can be effectively listed as:
0 , 1 , 00 , 01 , 10 , 11 , 000 , 001 , 010 , . . .
Obviously there is a machine that recognizes each of these by a halt.
But every sequence can continue in this set, so we would define every sequence strange again.
So the Solovay strangeness is a delicate duality of qualitative and quantitative finiteness.
The qualitative means that a finite machine has to be able to produce the beginnings and the quantitative means that the beginnings chance total has to be finite and thus narrow enough.
All obvious strangenesses that one could come up with, are Solovay strange with some well chosen machine.

So the negative of this, that is the sequences that are not Solovay strange for any machine are Solovay random. These sequences do not continue in, rather stop having beginnings from any Solovay strangeness. So we could call this the "Law Of Stopping".
Solovay strangeness has three equivalent but very different definitions.
I list them now shortly with giving their method of sequence collection and how an s binary sequence is confirmed to belong, that is "obeys" the strangeness.

1. Any effectively diminishing sequence of beginning sets.
   s obeys such sequence if all members have a beginning of which s is continuation of.
   Effectively diminishing means not only having diminishing chance totals but that the rate of diminishing is effective too.

2. A universal machine with some special features.
   s obeys such machine if it can compress the beginnings of s on the long run.

3. An effective betting sequence.
   s obeys such betting sequence if the payouts yield profit on the long run.

The first version was actually the new start of Randomness after the war by Martin Löf, a student of Kolmogorov. Solovay discovered the version that I regarded as basic only later.
Continuing a beginning from every member of the sequence may feel as too strong in Martin Löf's definition above in 1. Solovay's definition only requires that the sequence picks up infinite many from our single option set. But observe that skipping members, we still get a narrowing option set and so it's enough if we obey such partial option sequence.

We used this same idea when we proved the Raw Solovay Criteria for nil sets.
The equivalence of the effective versions is more involved and relies on Martin Löf requiring effectively diminishing not just mere effective diminishing. So diminishing sets that are effective is not enough. We need an effective rate of diminishing too!
When he created his definition his motivation was not a perfect correspondence with Solovay's simpler definition, since Solovay's came later. Instead, it was that the combined total strangeness could be defined by a single universal machine.

Four different definitions coinciding should mean that they grasped the objective single concept of Randomness. But here everything turned out "strangely" and differently as "expectable".
So Randomness splintered but we'll only show two other definitions beside the above four.
One widened the strangenesses while an other narrowed them.
The one that widened, simply modified the first above mentioned alternative road and thus required not effectively diminishing cover only effective diminishing cover.
Or in short, effective nil sets instead of effectively nil sets.
The sixth definition that narrowed what should be strange, came last historically too by Kurtz.
This is strange because this is the simplest definition, even simpler than Solovay's.
Also, it goes by claiming not strangenesses rather expectabilities, that is things that should be true for all random sequences. Of course the negatives then mean strangesses.
While the four equivalent definitions by strangenesses imply the Strong Law Of Large Numbers, this simplest by Kurtz's expectabilities is too weak for that.
The Kurtz expectabilities require that if we give an effective B collection of non continuing beginnings, with a total chance value of 1, then a random s sequence should be covered by the intervals. In other words, s must dip into our beginning set, that is a beginning of s must occur in it. So this could be called as the "Law Of Occurrence".
For example, if we claim a beginning with half chance then an other that is not continuation with a quarter chance, then a new one again with an eight chance and so on, then by Kurtz's Law Of Occurrence, one of our alternative predictions must be correct for a random sequence.
We might think for a second that this should be true for all sequences because among finite many options if they are excluding and have 1 total chance then one must come about.

But for infinite many options this is no longer true. Indeed, regard the following option set: $\{ 0 , 10 , 110 , 1110 , 11110 , . . . \}$. Its total chance is $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + . . . = 1$.
It is non continuing. And yet the sequence $111 \ldots$ will not have any beginning from our set.
In truth, such missing points from the 1 totaling disjoint intervals can be continuum many.
This is a third approach to the already mentioned Cantor Set.
These are the non random, that is strange sequences by Kurtz's Law Of Occurrence if the cut out intervals are given effectively. In fact, there is a fourth meaning too because these Cantor sets can also be called as zero sets. The point then is how these are different from nil sets.
Cantor sets are the leftover of intervals that total 1. Now if we go through these cut out intervals, that is regard more and more of them in any order but forming a single sequence, then the leftover of these always finite many intervals are again just finite many intervals.
But these always cover the final left over. So unlike a nil set that is just coverable by diminishing interval sets, these Cantor sets are coverable by diminishing finite many intervals.
This of course does not show any inner differences.
You'll find those in my mentioned book "Measure In The Unit Interval".
The fifth meaning behind the Cantor sets, or zero sets, is the most surprising.
Now these left over points from the 1 totaling disjoint intervals will almost appear "magically". To see this "magic trick" we have to use all left closed right open intervals.
As an initial step we sequence them in a single sequence and place them exactly in that order from left to right. They connect like beads and so we covered the whole $[0,1)$ left closed right open unit interval. Only one point is missing 1.
Now we can start juggling their order. The simplest trick is placing them in opposite order.
So we first place the left most to the right end. Thus 1 will be still left out.
Then place the second to the left of this one, then the third next to that, and so on.
The beads go backwards, but voila, now 0 disappeared!
The Cantor sets are nothing more than the disappearing points with super smart rearrangements.
As we already know, they can be even continuum many.

The Original Cantor Set

Cantor used an always fix one third carving out, so $p = \frac{2}{3}$ shrinking of the remainings.
A mini paradox is that no matter how small δ portion we use as fix cut out ratio, the remaining set still becomes a nil set. Simply because as we showed earlier the $p = 1 - \delta$ remaining ratio even though is almost 1 will still make p^n diminish.
Observe also that fix one third cut outs can easily be achieved by using base three infinite digitals like $.1010022100210222 \ldots$ and require that we don't use the digit 1.
The earlier mentioned glitch is here too, so the remaining sequences will not continue perfectly.
Of course, if we always cut out open intervals without the end points then we get a perfect remaining set but then we have to re-assign the binaries to this.
So this would be an alternative definition of the Cantor Space. An amazing difference is that the pulling apart makes a 1 size Cantor Space while this carving out method makes a 0 size one.
From inside we don't "see" this difference.
We can generalize our carving sequence as any kind of single interval as first level anywhere except at the ends, then two second level similar middle carvings and so on.
Is it then always true that the remaining set is not empty?
Is it then always true that it is a nil set?
As usual, the "no"-s and "yes"-s are "very no", and "very yes". And indeed here too, the remaining set has the same infinity as the full $[0,1]$ interval and can have any total length.
The reason for the first is that we can arrive to these final points through binary choices of the remaining beginning sections. For the second, that any desired total can be achieved by using carvings that approach that value. Of course there are tiny glitches.

For example, if we cut out not open intervals then the Cantor common point law is not applicable because the remaining final set is not closed. So then we have to steal the end points of the carved intervals temporarily and realize that they are only a sequence.
Finally a third question:
If we allow to carve out only points, can the remaining perforated set again be a nil set?
Remember that a nil set can be covered by arbitrary small totaling interval set.
And also observe that our carved out single points are a mere sequence of points.
And thus trivially a nil set because we can cover them by $\frac{\varepsilon}{2}$, $\frac{\varepsilon}{4}$, . . totaling ε.
So then if both the carvings and the carved out set were nil sets then their total the full [0,1] interval were a nil set. But the mentioned no under coverability of intervals refutes this.
All this was just an appetizer to see my other book titled "Measure In The Unit Interval".

<u>The n-champions and Kurtz's Law Of Occurrence</u>

Let the C(b) beginning feature mean having so many consecutive 1-s at the end of b that never occurs before in b. This is what we called earlier a champion b beginning.
We saw that by the occurring longer and longer 1 segments these champions must continue.
We also called as n-champion a champion that had n many 1-s.
Now a second argument can use the law of occurrence for each of these n-champions.

The 1-champions are: 1 , 01 , 001 , 0001 , . . .
The total is trivially $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$.
With two 1-s we have more possibilities: 11 , 011 , 0011 , 1011 , 00011 , . .
It's not visible at all why the chance total is 1.
But the fact that these are non continuing sets is trivial again.
The same will happen for longer champions and so if we can prove the 1 chance total in general then by Kurtz's Law Of Occurrence, the random sequences must bump into at least one from every possible set and so we get again that they all must occur in a random sequence.
Our proof of the 1 chance totals will be using some more general facts.
Let B be the set of those beginnings that contain less than n consecutive 1-s.
$\neg B$, the complement of B is all those beginnings that do contain n consecutive 1-s.
$(\neg B)^0$, the minimal subset of this contains all those beginnings that have no sub beginnings in $\neg B$. Thus in fact $(\neg B)^0$ is the n-champions.
Any one of the B^1 , B^2 , . . length groups of B plus $(\neg B)^0$ cover every sequence.
Indeed, a sequence either has some beginning in $\neg B$ and thus one in $(\neg B)^0$ too, or all in B and thus one in any B^m.
$(\neg B)^0$ is non continuing because a minimal subset is always such.
Also, a non continuing set has maximum 1 chance total. Both of these can be proven easily.
ΣB^m becomes smaller than any δ value. This would be a boring task to prove.
But more importantly, this implies that $(\neg B)^0$ has 1 chance total.
Indeed, enough to refute a $1 - v$ value and this would mean that with choosing δ as $\frac{v}{2}$, we could cover [0,1] with $1 - v + \frac{v}{2} = 1 - \frac{v}{2}$. This feels trivially impossible.
But the actual proof of this no under coverability of [0,1] is not trivial at all.
For these three proofs see the article: "Measure In The Unit Interval".

Proving the Zero-one Law

Enough to prove that if S is tail set and can be covered with p total then S can be covered with p^2 total. Indeed, then repeating the argument it can be covered with p^4, p^{16}, and so on. For $p < 1$ this means arbitrary small cover because if $p < 1$ then $p^n \to 0$.

To see this first, let $p = \frac{1}{1+\varepsilon}$. Then $\left(\frac{1}{1+\varepsilon}\right)^n$ will have in its denominator $1 + n\varepsilon + \, . \, .$

Now to show the coverability by p^2 let the covering sequence be $b_1, b_2, \, . \, . \, .$

If an s is covered by b_n then cutting this beginning off from s, the remaining sequence must be in S so it must start with b_1 or b_2 or . . So $b_n b_1$ or $b_n b_2$ or . . must cover s too. So the covering remains if we replace b_1 with $b_1 b_1, b_1 b_2, b_1 b_3, \, . \, .$ then b_2 with $b_2 b_1, b_2 b_2, b_2 b_3$ and so on! The total is $(b_1 + b_2 + b_3 + \, . \, .)(b_1 + b_2 + b_3 + \, . \, .) = p^2$.

Tail set's complement is tail set, so by what we mentioned about the non under coverability of [0,1] , only one of them can have 0 covering limit, that is be a nil set.
The theoretical possibility that both could have 1 covering limit, that is not coverable under 1 total, comes out of a paradox behind the zero-one law.
Let's call the set of sequences obtained by applying any beginning replacements to a single sequence a tail atom. These are the narrowest tail sets. Taking more and more together we get bigger and bigger tail sets. But they must jump from being nil sets, at once to become a 1 set.
Now comes a new level of randomness, namely deciding by a coin flip for each tail atom to be collected or not. Or in other words, randomly separate the set of all tail atoms. Then by obvious symmetry, we expect both halves to have same covering limits and then this can only be 1.
So we obtained the mentioned possibility of a tail set pair both having 1 covering limits.
Not with a concrete example, rather as a whole arsenal of plausible cases.
These may seem paradoxical in an other sense too if we realize that a random separation should mean half chances for both halves but regard the covering limits also as the expectable chances.
The error is of course in this second assumption. Coverings must have at least as big chance as the covered set. This is true, but a set and its complement can be so intermingled that the covering limits can not narrow down to the correct chance values of the sets.
Measurability describes this contradiction very well.
An S set is defined as measurable if for every ε there is some cover of S that the "over-cover", the points of the cover not in S and so being "wasted", can be covered under ε total.
This ingenious definition then implies that for disjoint measurable sets, their combined set can not be covered under the sum of the covering limits of the parts. But of course any combined cover covers the combined set and so the covering limits become additive. An other fact is that if $S \subseteq [0,1]$ is measurable then the $\overline{S} = [0,1] - S$ complement is measurable too.
Then by the previous additivity, $\overline{S}$ must have complementing that is 1 minus the covering limit of S. So our previous paradox simply meant that the random separations of the set of the tail atoms are not measurable. Of course, this wasn't an exact demonstration of non measurable sets again. Still staying with these random separation ideas we'll simplify our approach and forget tail sets and tail atoms. Instead assume a random separation of the sequences themselves, that is the points of [0,1]. Then the zero-one law doesn't apply and we may even think that now the correct half covering limits can come about for both sides.
But an other aspect of the random separations still suggest non measurability!
Namely, that both sides should penetrate all measurable sets with positive covering limit.
Then again both sides should have 1 covering limits. This was what Bernstein ingeniously created "concretely", using the Axiom Of Choice but without mentioning random separations.

Amazingly, an even earlier way of Vitali to exhibit non measurable sets went by something similar as the tail atoms. Remember, these are actually the replacement atoms.
Now using instead of beginning replacements only alterations, that is keeping their lengths and only changing the digits, we can define the beginning alterable sets and in particular the alteration atoms as alterations of a single sequence.
These are subsets of the tail atoms with a single tail which is not moved.
Amazingly, then the beginning alterations as beginning additions or subtractions will mean rational moves on the number line! So if we pick a single sequence from every alteration atom then this V choice set moved with every possible rational shift left or right and under 1 length will give all sequences and thus points of $[0,1]$. In fact, it will leave this unit interval too but still remain in $[-1, 2]$. Most importantly, the copies will be disjoint.
First of all, V can not be a nil set. Indeed, the rational shifts create a sequence of copies that cover $[0,1]$. But a sequence of nil sets is a nil set and so $[0,1]$ would have to be too.
So V must have positive covering limit and actually it can have arbitrary small positive ε value of this by simply choosing the points from a small enough sub interval of $[0,1]$.
Which of course doesn't mean to be a nil set because it's not one V choice set that has these, rather the different possible V ones.
More importantly, if V were measurable then its copies were too and these being disjoint would add together their covering limits. But then $[-1, 2]$ couldn't have a finite covering limit because infinite many copies are inside and $n\varepsilon$ grows above any value.

A strange Strangeness

The strangest twist in Randomness Theory is that the fifth randomness in our list that used simple effective nil sets became closely related to the third that used compressibility.
The biggest champion of this version was Chaitin. He was among those who fixed the mentioned glitch in Kolmogorov's idea. But he became most famous for his omega numbers.
And strangely, this caused the biggest doubt in his belief that all strangeness is compressibility.
The omega numbers are determined by a machine and form effective nil sets on their own.
They are not compressible, so are random by the second version. But only not diminishing effectively is that avoids them to be strange by Martin Löf's definition. So they are effective nil sets but not effectively nil sets and thus examples for the dilemma too, whether this wider strangeness concept is meaningful or not. If your head is spinning it's understandable.
How can something determined by a machine even be contemplated as random?
Is Chaitin crazy? Not at all! But observe that the word "determined" is actually undefined yet.
If I choose ten random numbers and assign them to ten machines one by one, then each machine will "determine" one of the random numbers. This of course is not determined by the operations of the machines. The omega numbers are! So observe the following deeper dilemma:
If this proper determination by a machine through its operation would be a precise concept, then sequences determined in this manner actually should be regarded as being Effective.
Then of course being random would be trivially insane.
So we have a more fundamental Effectivity challenge here.

Effectivity today is a collection of texts or natural numbers in particular.
So we could only talk about the places of the 0-s or 1-s as sets being effective.
Then the whole point of effectivity is that one of them can be effective while the other not. Effective sets can have non effective complement.
This became the new grand vision behind Gödel's undecidable statements, thank to Turing.
Then if we somehow establish that the theorems in a system is such effective set having non effective complement, the heuristic argument goes like this: Suppose that all statements and their negatives were pairs where at least one is a theorem! If there is a pair where both were, then of course we had a contradictory system. So assuming that this is not the case, in other words our system is consistent, then exactly one of the pairs is theorem. But then the non-theorems were effective too by simply collecting the theorems and then use formal negation!

Back to Turing, we have to admit that he didn't emphasize enough the one sidedness of his new effective collection method because the computing machine as means was more important at the time as the awakening of computers too.

Earlier, the dual effectivity, that is both halves of a collection and the complement being effective was regarded as the start. But it is a bad start!
One can only define some special ones among these, then get the full one sided effectivity and finally back again we get the full dual effectivity. In addition, functions were used as means, so the word recursive functions were the aimed dual effectivity.
Primitive recursive functions were the narrow class definable easily and then partial recursive functions were the means for the one sided effectivity. For collected objects then the recursive sets was kept for dual effectivity but for the one sided the idiotic term, recursively enumerable sets were coined!
Turing surpassed all this because his collection method at once collected these recursively enumerable sets. He should have thrown out this old name explicitly but instead his article's very title reaffirmed it in a sense. Indeed, it starts as "Computable numbers with . . . ".
We might think he is talking about natural numbers and collected effectively, but not at all!
He had further visions about real numbers and he meant these in the title with dual effectivity.
That is both the 0-s and 1-s being effective.
If only say the 0-s are effective then we only know effectively their places by a collection!
This wouldn't tell how soon a place is collected so the 1-s were not necessarily effective.
His usage thus was to call as computable the old dual effectivity, that is recursiveness.
All this wouldn't matter much but an insane and false idolization of Turing started much later.
In truth, it took time to sink in how fundamental Turing's new framework of Effectivity was.
These new parrots wanted to rewrite history! They claimed that Gödel, the absolute yardstick, at once recognized Turing's role which is an utter lie! But aside from this insane argument, even if it were the case it wouldn't make logical what they did. Namely, to rename the word recursive to computable but keep the main problem the word enumerable.
I repeat, the whole point is that unlike all other Effectivity approaches that widen some special dual effectivities to the full one sided, Turing at once defined this one sided effective collection.

Now back to our newer challenge, if such machine determination of a binary sequence is possible, then this indeed pokes a sharp question into the accepted effective collections.
In fact, it questions the so called Church or Church Turing thesis that claims a singular concept of Effectivity beyond the mathematically precise but only particular frameworks.
The status quo is of course that these new machine determined sequences and so the Chaitin omega numbers also, are not effective! Actually, the accepted term for them is recursively enumerable reals, or by the new computability translation, computably enumerable reals.
The strangest is that many mathematicians regard these as trivially not random, exactly by this determination but then simply are afraid to raise the effectivity problem!
The truth will come out! One day Effectivity will be challenged and will follow Randomness.
Today Randomness is defined by the completely artificial reliance on an imperfect Effectivity.
Back to the new machine determination, to make things clearer, I'll start with a hypothetical intermediate machine determination. Not yet relating to the omega numbers, neither to diminishing beginning sets. Then I will make the connection by showing machines toward both.
One that will use our hypothetical, to generate beginnings so can be used to define effective nil sets, and one that can produce our hypothetical from beginning generators.
Our hypothetical machine will use every n natural as input and will generate 0-s and 1-s.
For better visualization, imagine the naturals as a heading and under each, the generated 0 , 1 digits as columns. Now the big assumption is that we know each will fixate from a point.
So our machine will generate only 0-s or only 1-s in every column under a point.
As you could guess, these fixated digits will be the digits we use as our real number.
You can not argue with the perfect determination, only whether such machine exists at all.
But if so, then should this number we created be random or not?
Most would say not, and so defy Chaitin's faith in his vision.

Now first comes the second part, to use our column machine to collect beginnings because all strangeness definitions rely on this and not on the individual digits. So we do the following:
For every n value we let our machine go for all the $1, 2, \ldots, n$ inputs exactly n steps.
We ignore this initial running. Instead, we collect the n long beginnings that show up after, that is actually under. Infinite many are there but after finite more steps than n, all combinations must come about.
If none of the 0 , 1 values fixated in the n steps then we get all possible n long binary beginnings. But if some fixated then those will only appear with those digits.
The chance value of an n long beginning is $\frac{1}{2^n}$ and so if all 2^n many beginnings appear then the chance total is 1. But if one digit is fixed then the total is only $\frac{1}{2}$ because we have half of the cases. With more and more digits fixed the total chance is diminishing. Gradually, all the beginning digits are fixed too, so our beginning sets will cover only our fixated digits.
So our sequence as a single set is an effective nil set. But not an effectively nil set.
Because the diminishing was not effective. We had no bounds by n how small the chance totals are for the n group. We merely deducted the diminishing from the fixations.
It's interesting to see if the alternative Solovay strangeness fails too as it should. Infinite many of the groups can be combined into a single and our sequence must continue in this. This is okay but the total chance must be finite too. And this we can not guarantee without knowing how the group chances diminish.
Now we will show that our hypothetical fixating machine can exist.
Namely, can be manufactured from a simple beginning generator that generates a set with finite chance total. And such must exist because generating only finite many beginnings is such
So we merely attach a calculator that adds up the chances as the beginnings are generated.
Voila, the value of the total will fixate as more and more beginnings are generated.
In fact, the digits must fixate in such way that if one is fixated then all left ones are too.
A minor problem is that we don't know how big the whole part is in our infinite binary and so for a while the length of this whole part can change too.
A more readable binary total could be obtained if we knew that our total is definitely under 1.
If our machine generates beginnings that are never beginnings or continuations of an earlier generated one, then this stands. And so, such can be made from any machine by simply omitting newly generated beginnings that are beginnings or continuations of earliers.
For this of course we must store all earlier generations in our infinite memory.
Then we can only get a few digits after the decimal point by investigating our machine but to determine all digits is impossible for some machines. And this "some" can be made exact by using a universal machine as start. This universality is defined easier for not text generators rather text recognizers. Giving an input, the machine goes into work and halts if there is a result.
A universal machine is given the pair of an M machine's program and an I input as combined input. Then it will halt from this pair if and only if M would halt from I.
To duplicate a text and then let a universal machine use it as input thus means to ask if a machine would recognize its own program. Then it is evident that this duplication plus universal machine must share an input with every machine, that is halt or not in the same way.
But then this also means that this machine can not have a machine that is exact opposite of it, that is halts exactly where that doesn't. So we obtained a recognizable set of texts with non recognizable complement! To translate it into a similar result about generability we only must show that every recognizer is a generator too.
Feeding all possible texts as inputs can be easily sequenced but we wouldn't get through these because at the first negative try the recognizer machine will never halt.
This problem is easily overcome by the method of "dovetailing".
We go through the sequence of all possible texts as inputs but only process a few steps and then go to the next. After few steps we return to the earlier unfinished ones. Eventually, all those inputs will be recognized that would have been with working only on them.
So we can get a universal text generator too and then the complement, that is the non generated texts are non generable at all by any machine.

But knowing all digits of the chance total calculated for our machine would give the complement set and thus actually all complements of the generators.
These would also translate to a huge part of the unsolved problems of mathematics too!
So actually knowing this infinite decimal digit by digit would give a decision method for all these mathematical problems. Without proofs we would get verifications of them!
So the omega number name for this total as decimal is very appropriate.
But an alternate meaning behind this number is almost as exciting:

Instead of dovetailing, imagine a “timeless” possibility of all texts to be tried!
How could we guess the chance of a halt?
What jumps in mind is the already used $\frac{1}{2^n}$ chance of a single n long input.
We might at once reject this connection since the chance of all texts is $1 + 1 + \ . \ . = \infty$ by summing them in increasing lengths. But then we should remember that our recognizable inputs are non continuing and such set has maximum 1 total.

Usually it is less than 1 and this suggest that this sum should be the chance of halting or as it became called by Chaitin, the halting probability.
A more convincing argument is this:
Imagine a sequence of random coin trial sequences, that is random binaries under each other.
We let copies of our machine start to work on each, using the beginnings gradually.
Some may never halt some may halt. But every such halting, that is recognized n long beginning will be in average at every 2^n-th line present and disjointly, since by the non continuing every recognized beginning is only in one line.
Thus the halting portion calculated for more and more lines will approach indeed the sum.

<u>"Information" Paradox of Vitali sets</u>

Remember, that the Vitali set was a choice set from the alteration atoms of the binary sequences, thus points of $[0,1]$ and its rational disjoint shiftings produced the full $[0,1]$.
What we didn't mention there is that such Vitali choice set is perfectly possible for any alteration atom distribution of the evaluations on any E set not just the naturals. This paradox we'll approach now is universal too, though usually only the binary situation is mentioned.
The quotation mark in the title is meant to express a trivial mistake in the interpretations.
But if I'd say that it is "merely" a chance paradox, I would be incorrect too. It is true though that chance paradoxes are the deepest because we have stubborn plausibilities about chances.
The most obvious example of how we defy any mathematical sense in assuming chances, is the Rain Forecast Paradox: We hear in the morning news that rain has fifty percent chance.
Now suppose till noon it didn't rain! What will be the chance of rain in the afternoon?
Some might say less, because the forecast failed already in the first half day.
Some might say more because the full chance is now in effect for the second half.
Some would say it is the same. But then in any hour is it same too? This seems absurd again.
The official solution is that the question was meaningless. I don't quite accept this reply.
This paradox is handled similarly. And here I definitely reject the stupid avoidance by existing mathematical theories. Yet this is exactly what I encountered when hearing first time the $E=N$ version of it as the "Prisoners' Paradox" on a blog of Terence Tao about twenty years ago.
That's when I heard first time his name too and I remember how annoyed I was by his regurgitating reply about the Haar Measure to an idiotic blogger who pretended an "outrage" against this "bullshit Axiom Of Choice" that brings about such nonsense.
This prodigy since moved to America, got married and now refuses to reply idiotic questions.
So now the idiotic bloggers argue if Terence is the smartest person on the face of the Earth!
But let's forget the paradoxes of our "finite" world and see what this Prisoner's paradox was:
The 0 or 1 values we assign to the naturals are white or black hats placed on a sequence of prisoners by the evil warden. They can not see their own hat and they are not allowed to communicate. Now comes the cruelty, an execution of those that can not tell their own color.
The first thing that jumps in mind is that seeing the others should not help in telling their own color, so they all have half chance of dying. This then suggests that half of them will die.
So if we see a strategy with which only finite many will die, we are very surprised.
Then usually the strategy is rejected because it uses the Axiom Of Choice.
I will show that the strategy is perfect. There is nothing wrong with the Axiom Of Choice!
Instead, our tacit assumption was false that a half chanced strategy must imply a half death toll.
Or rather in reverse, a strategy with which only finite many of them die, making us jump to conclude that the strategy gave them an edge in deciding their own fate. So as if the information of the others' colors could help somehow through the illegitimate use of the Axiom Of Choice.
But actually, this most stupid interpretation of the paradox is self contradictory at once.
Indeed, the fact that only finite many will die follows only if they all follow the strategy.
If one prisoner follows it but the others don't trust it and so actually choose their colors spontaneously then it will not be true that only finite many dies. So the information of seeing the others' colors gives nothing at all! Only the others obeying the strategy gave an edge for all.
But if the others' colors giving an edge for a chosen one seemed unbelievable then the others' actions of choice giving an edge for a chosen one is even more absurd. So the simple truth is that there is no edge for an individual. The strategy leaves the half chance of dying still valid.
In fact, regarding any finite many of them as a group who all follow the strategy, gives also equal chance for any combination of their correctness or falsity.
So the strategy is not giving any edge for such finite group either.
Yet for any infinite many of them who all follow the strategy, only finite many of them can be wrong but these can be wrong with this total freedom of mistakes.
These remarks already dampen the paradox and indeed, all good paradoxes like all good magic tricks, are built by using more layers of deceptions.
Best example is the Achilles Paradox which combines the Finite Sum Paradox with an involvement of time and the false assumption that infinite many times must mean forever.

An obvious falsity because if the rain starts at noon then infinite many times before noon it was not raining yet. But the real trickery is to delude us even more by not using a motion toward a point when this abuse of time would become too transparent, rather one object chasing an other. So the faster Achilles who gave a head start to the half as fast opponent will reach the starting point of the slower opponent. Then of course the opponent will be ahead and when Achilles reaches this point he will be ahead again. This continues, so he will be ahead "always" and so "forever". In truth, he'll be ahead at infinite many places and times but these approach a point and time of passing.
In our present paradox the main extra layer is the assumption of a chance improvement.
To show its falsity most clearly from inside, we should use a bigger E than a sequence.
In fact, this way we get a much better paradox too with the extra layer of deception still present.
So let E be a screen. We use again black and white points, that is black and white pictures with infinite resolution as evaluations. Leonardo's Last Supper is then imagined as the perfect infinite though only black and white version. In this imagined universe then altering finite many points of this masterpiece means having finite many "errors".
An extreme picture is the all black screen and then the altered or errored ones are those that have only finite many white points. These can be night shots of an astronomer looking for stars.
Let's assume that our astronomer files these pictures in order of how many stars he saw.
So zero file is the single picture of the black sky. Next come all pictures with a single star.
Then with two and so on. Every point can be a star because this universe is infinite.
Now we fix a P point to be observed.
In the zero file there is no star, so P being a star has nil chance here for sure. In the next first file, from the continuum many picture only one has P as star, so the chance is again zero.
In the next we have all possibilities with two stars. That one of them should be P is a bit more likely than in the previous single star file but we still feel that the chance is zero.
And this continues because even a large number of assumed stars is a minute fraction in a continuum. So, in every file to find P as star has zero chance.
And thus overall too, to find in our filing cabinet any picture with P as star has zero chance.
The dubious step is of course, to pick a picture randomly from a sequence of increasing files.
Even to pick a file is questionable. Luckily this won't matter because now we do a re-filing:
We pull out from the single white point first file the one that has P as white.
We place this single picture after the zero, all black one. So we have these two singular pictures as start and then comes the single white file but with P being white missing from it.
Now we go into the original second file containing the two white point pictures. We pull out all those that contain P as one of them. These infinite many will be a new file placed again in front of the old. Then we pull out from the three star file all those that have P as star and these again will be a new file sandwiched in-between the two star one and the three star.
Amazingly, now every pair of files will contain exactly the same pictures, except P being black in the first then white in the second. Indeed, first on its own the all black and all black except P white pictures are such pair. Then comes the old single star file but P was pulled out so it is black in all. Then comes the pulled out two star pictures with one being P so indeed we got exactly the same as the previous single stars but with P white too.
And so on, we get perfect pairs, with first black P and then white P versions.
What this means is that P being black or white has exactly the same chance if we just pull out a picture from our file system randomly because now the questionability is paired too.
These two opposing results are a paradox in itself and somehow relate to our main one.
The main argument against the reliability of the re-filing result is that it was depending on P.
I don't regard this a fatal blow. After all, it is the chance of P that we ask for, so the point is already a selected one. An actual test could only be made by picking from the set of all possible evaluations and then reject those picks where we have more than finite many stars.
Unfortunately, the acceptable finite stars have zero chance, so we would have to wait forever.
Observe the most amazing part in this re-filing idea! Namely, that the size of E was irrelevant.
For Leonardo's Last Supper, the situation must only be modified as follows.

We make a green “originality” color added to the black and white pixels. When we alter these by exchanging white for black or in reverse, then we also alter green to red. Then we can file our finite alterations again by the green red colors. So we get again that in the finite alterations of the Last Supper we’ll have every P with same chance as being original or altered.
Now if we regard every black and white picture as a painting and by the Axiom Of Choice we pick one from every alteration atom as the original then the others are again just alterations.
A P point looking around can see the picture and identify the alteration atom.
If he chooses the “original” that we prefixed by the Axiom Of Choice, and he can also get the black or white value in this original at his own point, then he can use this as strategy.
This will work just as we explained above in any alteration atom.
So he will have half chance of being correct or wrong. In fact, any finite group with members using the same strategy will also have all correctness combinations with equal chances.
And yet any infinite group using the strategy will only be incorrect on a finite subset.
But now this full strategy version is not only important because it makes the final layer and thus produces the perfect seemingly singular paradox but also because it offers a return to the deeper paradox present in the singular alteration atoms. Namely, whether my belief in the half chance, that is no advantage interpretation of the strategy is valid or not. Indeed, now a test could be made by bringing about a sequence of full evaluations on E. In each applying the strategy, then we could see if indeed half of the time will our chosen P point be right or wrong.
In the prisoner version this means to imagine a sequence of alternate universes for our chosen prisoner. In each a different warden will live. For argument sake, being always evil but using different hat placements on the prisoners. I claim that in half of these universes our prisoner will die. So the strategy he used was superficial in the bigger picture of alternate universes.