Contents

1. Probabilities

An experiment where we don't measure results, merely observe if a chosen property will come true or not, should be called a trial for the chosen observed property.
The chance or probability for the success of a trial is a real number between 0 and 1.
In physical experiments, we don't have this p probability value of a property, and we do many experiments to find it out. In everyday trials, like throwing a dice, flipping a coin or drawing cards from a pack, quite on the contrary, we know the probabilities from the given physical conditions. In these situations, the major concern is how different properties relate to each other in their probabilities.
The most universal law is that if the success has p chance, then the failure has $1 - p$.
The logic behind this law is that a probability can be approached by the proportion of successes in repeated trials. So using n trials of the same property and finding k successful outcomes, $\frac{k}{n} \approx p$. Now if k was successful, then $n - k$ was failing, so the proportion of the failure is $\frac{n-k}{n} = 1 - \frac{k}{n} \approx 1 - p$.
The next basic law is that if a property $\mathbf{1}$ implies a $\mathbf{2}$ always, by simply physical reasons, then the probability of $\mathbf{2}$ must be at least as big as of $\mathbf{1}$:

$$\mathbf{1} \rightarrow \mathbf{2} \quad \Rightarrow \quad p(\mathbf{1}) \leq p(\mathbf{2})$$

The logic is that whenever in n trials $\mathbf{1}$ succeeds, then $\mathbf{2}$ automatically comes about too. So, $k(\mathbf{1}) \leq k(\mathbf{2})$.
From two $\mathbf{1}, \mathbf{2}$ properties, we can logically build the $\mathbf{1} \wedge \mathbf{2}$, $\mathbf{1} \vee \mathbf{2}$ properties.
$\mathbf{1} \wedge \mathbf{2}$ means $\mathbf{1}$ and $\mathbf{2}$ appearing together.
$\mathbf{1} \vee \mathbf{2}$ means at least one of $\mathbf{1}$ or $\mathbf{2}$ appearing.
Clearly: $(\mathbf{1} \wedge \mathbf{2}) \rightarrow \mathbf{1} \rightarrow (\mathbf{1} \vee \mathbf{2})$ and $(\mathbf{1} \wedge \mathbf{2}) \rightarrow \mathbf{2} \rightarrow (\mathbf{1} \vee \mathbf{2})$
Thus: $p(\mathbf{1} \wedge \mathbf{2}) \leq p(\mathbf{1}) \leq p(\mathbf{1} \vee \mathbf{2})$ and $p(\mathbf{1} \wedge \mathbf{2}) \leq p(\mathbf{2}) \leq p(\mathbf{1} \vee \mathbf{2})$.
So the individual probabilities are upper bounds for the "and" and lower for the "or".
Surprisingly, we can easily find an upper bound too, for the "or":

$$p(\mathbf{1} \vee \mathbf{2}) \leq p(\mathbf{1}) + p(\mathbf{2})$$

Indeed, the successes of $\mathbf{1} \vee \mathbf{2}$ can't be more than the sum of the successes:
$k(\mathbf{1} \vee \mathbf{2}) \leq k(\mathbf{1}) + p(\mathbf{2})$ This is so, because if $\mathbf{1}$ and $\mathbf{2}$ never succeed together, then exactly $k(\mathbf{1} \vee \mathbf{2}) = k(\mathbf{1}) + k(\mathbf{2})$, while if some trials for $\mathbf{1} \vee \mathbf{2}$ happen together, then it reduces this $k(\mathbf{1}) + k(\mathbf{2})$.
By analogy, we would expect that the lower bound for "and" should be the product, that is: $p(\mathbf{1})\, p(\mathbf{2}) \leq p(\mathbf{1} \wedge \mathbf{2})$
But this is not true. Indeed, if $\mathbf{1}$ and $\mathbf{2}$ exclude each other, then $\mathbf{1} \wedge \mathbf{2}$ never happens, that is, $k(\mathbf{1} \wedge \mathbf{2}) = 0$ so $p(\mathbf{1} \wedge \mathbf{2}) = 0$ too.
This asymmetry becomes even stranger, because the conditions at which the sums or products can be used to calculate "or" and "and" are also asymmetrical.
The exclusion that caused the falseness of "and"'s lower bound, is the condition that makes the "or"'s upper bound an exact value:

$$\mathbf{1}, \mathbf{2} \text{ exclude each other} \quad \Rightarrow \quad p(\mathbf{1} \vee \mathbf{2}) = p(\mathbf{1}) + p(\mathbf{2})$$

The logic was inherent already in how we obtained $\leq$.
Indeed, if $\mathbf{1}, \mathbf{2}$ never happens together, then $k(\mathbf{1} \vee \mathbf{2}) = k(\mathbf{1}) + k(\mathbf{2})$.

Throwing a 6 on a dice has $\frac{1}{6}$ chance. A 5 has the same $\frac{1}{6}$. Throwing 5 or 6 has $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ because these two exclude each other.
But throwing two dices and looking for 6 or 6, which means $\mathbf{1} \vee \mathbf{2}$, with $\mathbf{1} = 6$ on first dice , $\mathbf{2} = 6$ on the second dice, the claim $p\,(\mathbf{1} \vee \mathbf{2}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ is false. The chance must be a little less, because sometimes they are both 6.
The crucial new condition for calculating the "and"-s is <u>independence</u>. This means that **1** and **2** don't influence each other at all. This actually means that among the **1** outcomes, the **2** has its own p (**2**) chance and among the **2** outcomes, the **1** has its p (**1**) proportion too. So:

$$\mathbf{1}, \mathbf{2} \text{ are independent} \quad \Rightarrow \quad p\,(\mathbf{1} \wedge \mathbf{2}) = p\,(\mathbf{1})\, p\,(\mathbf{2})$$

Indeed, $$k\,(\mathbf{1} \wedge \mathbf{2}) \approx k\,(\mathbf{1})\, \underbrace{p\,(\mathbf{2})}_{\frac{k\,(2)}{n}} \approx k\,(\mathbf{2})\, \underbrace{p\,(\mathbf{1})}_{\frac{k\,(1)}{n}} \approx \frac{k\,(1)\, k\,(2)}{n}$$

So, $$p\,(\mathbf{1} \wedge \mathbf{2}) \approx \frac{k\,(1 \wedge 2)}{n} = \frac{k\,(1)}{n}\,\frac{k\,(2)}{n} \approx p\,(\mathbf{1})\, p\,(\mathbf{2})$$
The most typical independence is the repetition of an experiment or trial.
So throwing two 6 with two dices has $\frac{1}{6}\,\frac{1}{6} = \frac{1}{36}$.
Most amazingly, this product law allows us to calculate the "or"-s even when they are not independent and so the sum would be false. The "trick" is to use the complements, that is failings instead of successes. Indeed, for example throwing at least one 6 with two dices can be looked first negatively, that is having no 6 at all. Neither on the first dice <u>and</u> neither on the second. The chance of this is $\frac{5}{6}\,\frac{5}{6} = \frac{25}{36}$.
Then of course, the "at least one 6" is the complement of this, that is:
$1 - \frac{25}{36} = \frac{11}{36}$ which is indeed less than $\frac{1}{3}$, as it should be.
A more meticulous and universal method of getting the right probabilities is the "case counting". The cases are the individual possible outcomes. We simply have to list these, go through them and count the desired outcomes as k, to obtain $\frac{k}{n} = p$.
For example, flipping two coins, the cases are:
(head , head) , (head , tail) , (tail , head) , (tail , tail) so $n = 4$.
Getting two heads is one case, so has $\frac{1}{4}$ chance, as it should be by : $\frac{1}{2}\,\frac{1}{2} = \frac{1}{4}$.
Having at least one head has three cases, so has the chance of $\frac{3}{4}$ not $\frac{1}{2} + \frac{1}{2} = 1$, which would be absurd. So, now we see that the earlier mistake with dice for at least two 6-s as $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ would be just as absurd if we go on and use six dices.
Then getting one 6 out of the six dices would be $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1$.
But throwing six dices doesn't guarantee a sure 6.

For two dices the cases are:

(1 , 1) , (1 , 2) , ... , (2 , 1) , (2 , 2) , ... , (6 , 6). This is 36 cases.

The ones that give at least one 6 are: (6 , 1) , (6 , 2) , (6 , 3) , (6 , 4) , (6 , 5) ,

(6 , 6) , (1 , 6) , (2 , 6) , (3 , 6) , (4 , 6) , (5 , 6) . This is 11 cases.

So indeed, the chance is $\frac{11}{36}$. With this case counting system, the complementarity is merely a trick to use less calculation. Indeed:

What's the chance of getting at least one ace if we pull two cards from a pack of 52? The total number of cases for the pulled two cards are $52 \bullet 51$ because the first card can be 52 kind, the second only 51. The desired cases are either the first pulled being an ace and thus, $4 \bullet 48$ many, because there are 4 aces, and 48 non aces. Or in reverse, the second pulled is an ace only, which has $48 \bullet 4$ cases. Or, both being aces, which has $4 \bullet 3$ cases. Thus, the derived probability is:

$$\frac{4\bullet 48+48\bullet 4+4\bullet 3}{52\bullet 51} = \frac{4\,(48+48+3)}{52\bullet 51} = \frac{4\bullet 99}{52\bullet 51}$$

Using complements, first we go for not pulling ace at all. This has $48 \bullet 47$ cases. So the desired chance is:

$$1-\frac{48\bullet 47}{52\bullet 51} = \frac{52\bullet 51 - 48\bullet 47}{52\bullet 51} = \frac{4\,(13\bullet 51-12\bullet 47)}{52\bullet 51} =$$

$$\frac{4\,(13\bullet 47+13\bullet 4-12\bullet 47)}{52\bullet 51} = \frac{4\,(47+13\bullet 4)}{52\bullet 51} = \frac{4\bullet 99}{52\bullet 51}$$

This only seemed long because I wanted to show the same result.

Pulling say five cards, the complement method remains the same, but counting the successes is very long.

Beside the obvious mistake that can happen in the case counting, namely miscounting, a more hidden one is if we use cases, that are not really the elementary equal chanced possibilities. A most brutal example could be for throwing two coins, using as cases:

(both heads) , (both tails) , (they are different)

This is a false "casing", because the (they are different) has more chance than the other cases. Indeed, it hides two actual cases.

The cases we had in mind up until now were collections of possible "flows" or physical processes. Throwing a 6 with the dice as case is the set of all possible droppings and bouncings that end in a 6 on the top. Sometimes we have to go deeper than the cases, in other words, we must regard the flows themselves as cases.

Throwing a dart onto a board, we can use conventional cases as collections of flows that land in the different rings, but to calculate the actual chances, we must use areas. So it's better to go deeper and use the individual flows as cases. The landing in different rings, then are not cases, rather events, combined from infinite many cases.

So probabilities can be built on Set Theory. Namely, the measures of sets will be crucial, so Topology will be the foundation of such theory. This line was axiomatized by Kolmogorov, but as straightforward it seems, it was actually an escape from something very fundamental, namely randomness.

2. Randomness

The crucial fact that probabilities for "or" don't add up for independent trials, means that throwing six dices, we don't have $6\,\frac{1}{6} = 1$ sure success. This is obvious. We can throw even a million dices, and get no 6 at all. The theory of chances is perfect. It claims addition of chances only, if they are exclusive. Of course, the very assumption of each dice landing independently, means that they are not exclusive. They can happen in any combinations, even all million landing on 6. This of course would happen extremely rarely, namely with a chance of only $\frac{1}{6}$ to the million power.

To find only 1 , 2 , 3 , 4 , 5 landings on all the million dices but none 6 , has a chance of $\frac{5}{6}$ to the million power. This is much bigger than the chance of all 6. But both $\frac{1}{6}$ and $\frac{5}{6}$ to the million power are so minute that we simply can not distinguish them subjectively, they both seem to be almost zero.

The fault of the probably theory comes in when we go to infinity.

Already instead of a million dice, it's better to imagine to throw one dice a million times repeatedly. For infinite many dices even more so, so we envision an infinite trial sequence and their outcomes:

s = 3 1 6 2 1 4 2 6 3 6 4 5 1 6 5 . . .

Of course as we said, usually we watch out for a particular event like the six at dice, so a better sequence is using 1 for these successful outcomes and 0 for all others.

So for the above dice outcome sequence we would have:

s = 0 0 1 0 0 0 0 1 0 1 0 0 0 1 0 . . .

Our intuitions are correctly the same for infinite many trials in space or in time.

So, consecutive trials is merely a convenience in language. But there is a hidden trap in this convenience! Indeed, in space we can imagine infinite many trials spread out without an order. Such discrete point sets can always be sequenced by taking bigger and bigger volumes around a chosen fix origin and take the newer and newer finite many points. But this sequencing idea is not obvious at all! It's the result of Set Theory! Trying to use Set Theory even further and imagine abstract set of trials is however out right contradictory, exactly by Set Theory! Indeed, infinite sets can be arbitrarily stretched and the ratios of successes or failures can be altered. For example, the successful six throws as an infinite set can be easily doubled by sequencing them and then replace the n-th with the 2n-th and use the thus freed up odd ordered ones as new set of successes. In short, chances can only be meaningful in structures not in pure sets. A sequence is the simplest infinite structure but isn't using this a doomed simplification? We ignore this question now because there are bigger problems.

First, the crucial point why probability theory is failing when we go from finite to infinite many trials:

At finite many trials, all mathematically possible outcome combinations are possible!

In fact the mathematical probability calculations can tell with what chances such combinations appear. These calculations are not so hard to generalize to infinite many combinations and then many times 0 or oppositely 1 chance value can come out. This in itself should not mean impossibility or certainty, especially because already at continuous flows we had 0 chance for a single flow. To hit a particular point on a dart board has 0 probability yet it is possible, in fact every outcome is such. Of course a known or chosen point to be hit "again" seems to be truly impossible, so there is hidden deep problem of prior knowledge or information. This was ignored by

probability theory. But with non continuous 0 limit results the possibility or impossibility of 0 chance became much more concrete. Namely:
Certain 0 chance combinations are felt instinctively impossible by everybody!
This healthy negative approach unfortunately at once turned into some positive claim when we wanted to be more specific.
A million dice throws without a single 6 is rare but possible!
Infinite many throws never landing on 6 is impossible!
Not only must 6 happen but it must happen in $\frac{1}{6}$ proportion in the long run.
Observe, that logically, the claim that trying infinite many times, anything that has any chance, must happen, already implies infinite many occurrences. Indeed, after one a next must happen, then again and again. So, the infinity of 6 occurrence follows from the single occurrence. Something very deep is lurking in this re-occurrence argument, which was ignored in the chase for the even stronger claim than the infinite occurrence. This stronger claim was the above mentioned fact that the limit proportions are the probabilities. It was even called the Law Of Large Numbers.
The obsession with this has its justification! After all, it is the probabilities that the theory of chances must assume given by physics! So reducing them to limits of outcome ratios seems to bring about a selfcontainedness of mathematics.
But this is a false vision! There shouldn't be selfcontainedness, in fact it has to connect to an other "external physicality", Effectivity!
The simplest "failure" of the mere law of outcome proportions is the following throw sequence of a dice: $1\ 2\ 3\ 4\ 5\ 6\ 1\ 2\ 3\ 4\ 5\ 6\ \ldots$
The six possible sides appear in tidy sixth proportions but it's too tidy!
The heuristic idea to rescue the Law Of Large Numbers as the lord of randomness, was to introduce the concept of the "observational sequence".
Instead of the full sequence, we can choose for example to regard only every second or sixth outcomes and in that sequence the Law Of Large Numbers must also remain.
This of course at once excludes same outcomes in predetermined places as above.
But requiring that all subsequences have the same proportional limits, opened up a new can of worms. First of all, just to use fix positioned subsequences is still too narrow! For example, if we see that in a dice throw sequence after every $2\ 3$ outcome doubles a definite 6 follows, then this couldn't happen randomly. We might say "no problem", use the next positions after the $2\ 3$ doubles as the observational sequence. If all must go in this subsequence by the Law Of Large Numbers, then all six is impossible again. In other words, we should allow observational sequences determined by not just fix positions, rather by the outcomes themselves too! But this has an obvious contradiction in it! Namely, we can regard as observational sequence the 6 outcomed positions themselves. Then of course it must be all 6. We seemingly can dodge this contradiction by observing that in our previous example the observational places depended only on the last two, that is earlier outcomes, not on the outcome of the selected position itself. So a seemingly perfect general solution is to allow observational sequences as next positions to certain beginnings.
But now a really deep problem becomes what should "certain" mean. If any selection of beginnings can be "certain" then we can also regard the beginnings before the 6 outcomes as certain and thus we are back to square one, having all 6. So the "certain" beginnings must be chosen without a "knowledge" of the next outcomes.
That's when Effectivity entered the scene because Church, one of the persons who was obsessed with randomness happened to be one of the fathers of effectivity too.
So it's like a Greek tragedy with many possible fathers and lots of obsessions.

But most importantly, Church believed in Effectivity much deeper than his own approach to define this still mysterious entity. Now, if there is an objective absolute reality in Effectivity, then choosing the beginnings effectively, not only excludes the “knowing” of the next outcomes, but it puts randomness onto an almost perfect platform, a platform between math and physics. The tragedy was, that in spite of all this grandeur, the positive, that is Law Of Large Numbers approach to randomness was wrong! So a castle was built on false foundation.

Randomness is not about what must happen in infinity, rather what must not happen.

In other words the real foundation are “strangenesses”. These are the particular infinities and in a random sequence all such must stop, fail after a finite occurrence.

In fact, even the Law Of Large Numbers is only seemingly about the infinite tendencies. It actually claims the stopping of some strangenesses! I’ll explain this in detail but now I want to show it as it appeared to me originally.

This kind of detour never happened in a math book before and I have to admit that I also only added it on a second rewriting. I rewrote these pages not with this goal, rather because reading them I realized I can express the points much simpler.

But then the truth just brought some honesty with it too.

I wasn’t good in math in elementary school. I wasn’t good in anything! I applied to the planned high school that was supposed to open in my elementary school. These “upgradings” were necessary because a boom generation born after the war. But the Ministry of Education still had some requirements and “thanks God” my school was refused the permit. So all students had to look for other places in the summer three months. I definitely didn’t want to go to the boy high school my older brother just finished. That was the closest and next to it there was a girls one that tried to change to mixed classes but in that year only for a singe class with special math curriculum.

There were two famous special math high schools in Budapest already and this passionate but unknown math teacher in a girl’s school only got the opportunity to open a new one because the principal, also a math teacher who never actually taught math, was well positioned politically. Of course I didn’t know all this when I went to apply, purely to my mother’s persuasion. There was no entry test only a conversation but I was sure I failed even that. As it turned out, they got the permit also in the last moments, very few students applied and so I was accepted.

I have to add a little but I think very vital detail. My mother’s insistence hadn’t been enough to make me overcome my deep disbelief in myself, without an incident that happened a year earlier. My regular math teacher, a psychotic woman was sick and a substitute teacher named Kozalik came in. I don’t remember what happened exactly but he some questions, I went to the blackboard, he asked me what grade I had and he remarked that I could be much better. He said: You should start to believe in yourself and open the textbook in that mood. This stuck in my mind and came back when my mother suggested the seemingly absurd idea of me going to a math high school.

So there I was! On our first math class the teacher Laszlo Banhegyi gave a big speech, how mathematics is not about how much you know and how a child can solve the biggest problem and how we all got a new start there. As an example he mentioned this famous professor who in Moscow was asked a question that he couldn’t answer yet an elementary school kid solved it at once. The question was to tell the resulting formula of a product from the members $a-x$, $b-x$, $c-x$, and so on, for all letters of the alphabet. I put my hand up and said that the result is zero because the last letter of the alphabet is x. Of course the last letter of the Hungarian alphabet is also z, so I was wrong. But I got two highest marks instantly, regardless of my mistake. Indeed, x is in the alphabet before z and y and so $x-x$ does make the product zero. My teacher didn’t realize that my mistake was not my momentary mix up of the

alphabet rather a wishful blindness. I thought that a product becomes zero only if the last multiplication makes it so. I didn't know that algebraically, variable products can be calculated and this represented the woods from which the forest may not be seen.
I really didn't know any math!
Six months later I won the Hungarian high school math journal competition and by that time I did know a lot of math.
Across the high school there was a little church. On that first day, after school I went there and sit alone for almost an hour. My communist parents didn't know this just as I didn't know that I was Jewish. It was between me and God.
The fact that things can come to you from the air, and you can know things that you have no idea about, was never a question for me. Yet probably I denied it to even myself. There were three distinctive such mathematical visions that came to me without any readings or exposures about them.
The first was that choosing infinite sequences of numbers to be in sets can "force" properties of the possible sets. I had no idea why this can be important and the word "force" came to me in Hungarian since I didn't speak a word English.
The other more definite vision was to extend the continuous line by putting into the pints, lines themselves. Here, I saw a Connection that this allows to talk about limits as new infinitely small actual numbers. I started to develop a system and sent a letter to Werner Heisenberg to suggest that such new numbers could be used in Quantum Mechanics. He replied in few days!
The third most definite vision was about randomness! Here I had a very concrete starting point. I found it amazing that in an infinite sequence of coin tossings, repetitions or opposite repetitions of a beginning must stop.
So 0 0 1 repeating in 0 0 1 0 0 1 is not strange yet, nor oppositely in digit sense, that is as 0 0 1 1 1 0 or in space sense as mirrored 0 0 1 1 0 0 or in combined sense as 0 0 1 0 1 1. But such repetitions or mirroredness of the beginnings become more and more unlikely as the beginnings themselves grow and must stop.
I went around in my family and asked everybody's gut feelings and was very happy that they all had the correct a priory "knowledge" about randomness. But what puzzled me most, was realizing that this stopping of an unlikeliness is hiding a complicated rule. It's not simply a diminishing chance that causes the gut feeling! Indeed, we can create beginning properties that tend to zero in chance for bigger and bigger beginnings, yet we don't feel that they have to stop. The simplest such construction can go like this: Longer and longer runs of full 0-s or 1-s like heads or tails in a coin tossing sequence must occur! To see this is easy by regarding the throw sequence as repeats of any fix "windows" like say thousand length mini throwing sequences. Since we have infinite of these, thus every thousand long combination must appear including full heads or tails. And since the window can be arbitrary big, thus we have arbitrary long repeats. Now, in a beginning we can check how long is the last outcome's repetition. For example, in 0 0 1 0 1 0 0 this is 2 , in 0 1 0 0 1 1 0 it is 1, in 1 1 0 1 1 1 1 it is 4 and so on. If all the repetitions in a beginning are shorter than this last then it means we have a beginning with a new repeatition champion. The chance of a beginning to have such new champion is getting less and less as the length grows. Yet, since there are arbitrary long repeats, there have to be infinite many such champion creating beginnings too, namely the beginnings containing the longer and longer repeats at their end.
So why do the beginning repeats or mirrored beginnings have to stop while other diminishing chances can occur infinitely? The natural answer is that there has to be some distinction among diminishings. It doesn't take long to realize the concept of fast versus slow diminishing. But there is something here beyond this fairly easy

distinction! Namely, instead of particular properties we can ask how particular sequences can possess these fast and slow diminishing properties. If for example we knew exactly by a rule when beginnings having new repetition champion must appear, then we could turn that into a new fast diminishing property and it would indeed be impossible for a random sequence to exhibit infinite many times. In fact, even the question of an effective super sequence without obeying any fast diminishing property infinitely, occurred to me. So I had a vague feeling that randomness can not be absolute. The different diminishings or zeros meant a relation to the other fixation, the deep ordering of the line too. But I saw no lights in these tunnels. Looking back now, the most surprising is that the concept of Effectivity did not occur to me at all. But there is more to this blindness because in grade three and four we started computer studies. Then when I went to America I was at the birth of the small computers. In fact in Portland at the Tektronix company I made suggestions to use the oscilloscope monitors for the computers and incorporate the IBM keyboard into the language. I couldn't envision the "home computer" yet but I did envision the "office computer". When I was neglected, it made me a good excuse to leave the whole god damn industry and return to pure math. This happened with big detours. But then decades passed by and I still missed the whole Effectivity or "computability" band wagon.

Back then in high school, I also saw a connection of a certain mirrored beginning stop with the simplest unsolved problem, the Goldbach conjecture. This claims that every even number is the sum of two primes. But this means that every number is the average of two primes. In other words every number has a prime under and above same distance away. So if we place 0-s for composites and 1-s for primes, then in this sequence, mirroring a beginning will always make two 1-s collide. The falsity would mean that there is a beginning that when mirrored doesn't collide at any 1 value. So it is anti symmetrical in this special sense. This is a definite unlikeliness, so it should stop if the primes are random. The primes are not random but this unlikeliness never occurs.

Much later I learnt about these three things in the same order. Cohen's book on the Continuum Hypothesis was my first English book in Rome and I decided to see him.

I read about Non Standard Analysis in the Stanford Math Library when I worked there waiting for Cohen to return from England.

I only started to read about Randomness here in Australia fifteen years ago. And this is the subject that I must return now.

A most heuristic though very vague claim about random sequences is that we can not claim anything about them! In truth, two far from obvious exemptions must be made.

One is finiteness and one is naturalness.

As a random sequence unfolds we can see how it goes and can tell these beginning features. These are then finite ones. But we can tell even less, namely conditionally that after certain beginning certain segments will or can not follow. These still are beginnings, so actually we claim or exclude some beginning combinations.

We can also "predict" infinite features but these then are "natural" that is true for all random sequences.

So, any valid $P(s)$ claim about an s random sequence is merely a combination of beginning features or natural ones, true for all random sequences. The combination itself of course is excluding some random sequences due to its finite parts.

This fact that the individual s random sequences can only obey combinations of finite or natural claims is a good start but it doesn't lead to a definition of randomness.

Also observe that we have an arsenal of non random sequences, created by us and these again obey two kinds of claims, finite beginning ones that are exactly the same as for a random and infinite ones.

These infinite ones can be same natural or boring features that are true for all the random sequences but some can be unnatural, not true for any random sequence.
Combined features again mix up and we can claim "or" combinations too that widen the scope of our claims. The above usage of "unnatural" was infinite property that excludes all random sequences. We might exclude all random sequences but include some finite claims and then we use the expression "strangeness".
To have infinite many 0-s or 1-s in a sequence are both natural.
To have only finite many 0-s is unnatural.
To have all 0-s is a strangeness because it tells about beginnings too.
In general, P(s) features about sequences can really be confusing.
Finiteness mixing with infinity is the main reason and we'll solve this problem first.
Any finite sequence can come about randomly, so can be the beginning of a random sequence. Also, a beginning shouldn't influence the continuation, so combined:
A random sequence remains random if only its beginning is changed.
Such beginning change is called alteration and so the set of random sequences is invariant for alterations or in short: alterable.
A beautiful feature of this alterability is that if a P property has it then its opposite $\neg$P has it too. Indeed if an altered s sequence would become P from $\neg$P then the "back altering" would contradict the staying in P that is the alterability of P itself.
So we have alterable pairs and these are the crucial key for a promising definition of randomness. Alterability is the exclusion of the finite claims that is an exactification of our "infinite" above.
An alterable property still can be a boring naturality, true for all random sequences but can also be unnatural.
This then is true for both members of the pair and so we could imagine that both P and $\neg$P are natural or both are unnatural .
Now comes the big claim: This is impossible!!!

From a pair of alterable properties one is always natural the other unnatural.
One contains all the random sequences the other none!

If this is true then the only problem would be to tell for each alterable P which is the natural and which is the unnatural among P and $\neg$P.
One way would be to find a single random s in one of them because then it is the natural. But if we use an axiom system to define these P properties then we would prefer a formal way of choosing too.
Instead of using axiom systems we can regard the P properties as collections.
Surprisingly, the abstract concept of sets would not work.

Indeed, lets pick a random sequence s and regard the S set of sequences obtained from s and adding all of its beginning altered versions as well.
So S contains all sequences that end like s and only these s ending ones.
Clearly S is alterable but it can't be usual because it doesn't contain all random sequences only the s ending ones. But it can't be unusual either because it does contain random sequence, namely s. In fact it only contains random sequences namely all the s ending ones!

Observe that the previous axiomatic and thus formal or explicit collection of sequences as P properties avoided this contradiction because there we can not start with a random s.

An other direction of collection can avoid this too.
Instead of explicit we use effective that is finitely determined collections.
A finite system like a machine can collect infinite many objects, but the objects themselves must be also finite. So sequences can not be used maximum segments.
Then two possibilities arise. We combine such segments after each other to define the sequences or we combine them all being just beginnings.
The first is clearer because the segments follow independently but we have to specify what consecutive lengths should be allowed. With beginnings, the widening, the continuation is enough but this restricts the choices themselves.
This beginning determination became the successful way and it dictated that the whole obsession with infinity that is alterability should be given up.
The aim is not to find all the unnatural claims rather some wide class of strangenesses. If our class of strangenesses can eliminate all the random sequences then we will still end up with an easy definition of randomness as the sequences not obeying our particular strangenesses. The more concrete practical verification is that for any suggested and reasonably acceptable strangeness we have to show how our class of machine generated strangeness can imitate that strangeness.
First of all, observe that to turn a strangeness into unnaturality is quite easy in fact any property can be infinitized that is turned into alterable by simply adding to it the condition "from a point".

The solution to define some P(s) properties is to compromise! We don't want to grasp all intuitively meaningful properties only some very special ones. Namely, the ones that claim about an s sequence that certain beginnings are appearing finite or infinite many times. So as P(s) we only regard the following ones:
There are only finite many b beginnings of s so that B(b). Or oppositely:
There are infinite many b beginnings of s so that B(b).
These opposing pairs are not beginning independent! Indeed, changing a beginning changes all beginnings and the infinity can easily turn into finiteness or vice versa.
You might even wonder why the hell I went on about the heuristic split of the end properties into universality or strangeness if these won't be usable for the properties we'll use. The answer lies in the already mentioned non pure non end strangenesses.
Above I said that they contain finite "dirt".
These impure strangenesses are our salvation! These finite impurities allow the selection of special strangenesses that already exclude and thus define randomness.
Only B(b) properties became successful, not some theoretical P(s) ones.
Actually, the B(b) property is regarded as B collection. In short, B(b) is $b \in B$.
In fact, the set of all possible beginnings is an easily listable set as :
0 , 1 , 00 , 01 , 10 , 11 , 000 , 001 , 010 , 011 , 100 , 101 , 110 , 111 , 0000 , . . .
So B is merely a selection of elements from this list.
Such selection can be done by a machine and so we actually regard all possible machines as the B beginning collectors. To turn actual B(b) beginning properties into B machines is not that easy sometimes but can always be done.
As I mentioned, Church used this idea first to select beginnings, so that the next positions in an s sequence could be used as observational sequence to claim the Law Of Large Numbers. Now we use them directly as collections that must appear finite or infinite many times in s. But what should be the magic wand that decides whether B can or must contain finite or infinite many beginnings in a random s ?
Here comes in a second incredibly simple feature of these B collections!
Namely, every b element in a B has a well determined chance value.

For coin flips, this merely depends on the length of a b because the particular combinations of same length are all equally possible half chance multiples.
For example, 0 1 1 0 1 is just as possible as 1 0 1 0 0.

They both contain five half chance choices and thus have a chance value of $\left(\frac{1}{2}\right)^5$.

For dice throwings with 1 being a six while 0 anything else, the chance values are still fairly simple but they depend on the 0 1 combinations too.

For example, 0 1 1 0 1 has the chance: $\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^3$, while 1 0 1 0 0 has: $\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$.

Using these chance values in a B collection is a very barbaric idea though!
Indeed, these chance values are only meaningful as chances among the same length trials. So for a B collection, the totals are not a full 1 chance value, instead can be arbitrary big even infinite! In fact, infinity is more likely and finiteness is a sign of smallness, that is B being a rare collection. This guaranties already that we have a strangeness, random sequences can not have infinite many beginnings from B.
The reverse is obviously not true as I will show it in a second but this is not a real problem. We don't aim for a total system of strangenesses anyway since we only use beginning properties and their infinities. So even if some infinite chance totaling B sets are strange, we can still hope that the finite total ones are sufficient to exclude all non random sequences. In fact, these trivial examples for non reversal of the finiteness also support the finite total as the good concept. Most importantly though, we can easily show why finite chance total of B must exclude randomness with two heuristic steps. One is a simple mathematical consequence of the finite total, the other is regarding a particular zero chance as impossibility.
Unfortunately, later there will be deeper doubts whether this finite chance total of B is an absolute concept. So there will be alternative wider notions of strangenesses that therefore can only be the infinity from an infinite total chanced B. But these doubts are on a higher level, not meaning any deficiency in the finite chance total of B.
I mentioned that in high school I regarded the strange beginnings, and also how Church tried to use them. Now finally the B sets are our solutions.
But for an outsider this whole obsession with beginnings is not as natural. To predict individual 0 or 1 is the real start to defy randomness. And then, if we envision such one by one prediction to be very rare or actually rarer and rarer, that is at positions further and further ahead, then this feels like very minor restriction on the beginnings.
A simple elementary school paradox can bring us closer to reality! Lets eliminate all the numbers up to one million that contain a chosen digit say 9. What percentage of numbers remain? The shocking answer is less than 50 percent. Only checking the percentages gradually, will show what's going on. Up to ten we have ten percent, namely the 9. Up to hundred we have nineteen percent because ten percent is the nineties and in the remaining ninety percent we have ten percent as the ones in every ten group. Up to thousand we'll have twenty seven percent namely from the ten percent ninehundreeds, the nine percent nineties among the rest and finally the ten percent of the remaining eighty one percent is about eight percent. As we see the formula is simple, ten percent plus ten percent of the rest, plus ten percent of the rest and so on. And this adds up to large percents very soon. The individual outcome restrictions similarly add up to very large beginning exclusions, that is very small possible beginning chances. But here as we select from the set of all beginnings, the non predicted lengths can be included in full, thus increasing the total B extremely.

Of course a more careful selection can exclude those that will be narrowed anyway and then we indeed get a finite total!
For example lets go ahead fast in our restrictions of say 0 outcomes, namely by the powers of 2 , that is at the places: 2 , 4 , 8 , 16 and so on:

_ 0 _ 0 _ _ _ 0 _ _ _ _ _ _ _ _ 0 _ _ _ _ _ _ _

The underlined positions can be 0 or 1 and thus a full beginning list with their chances under them is:

$$\begin{array}{cccccccccccccccc} 0 , & 1 , & 00 , & 10 , & 000 , & 001 , & 100 , & 101 , & 0000 , & 0010 , & 1000 , & 1010 , & \ldots \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \ldots \end{array}$$

As we see, the total will be definitely infinite because we have groups with 1 totals.
Now lets keep only the restricted lengths:

$$\begin{array}{ccccccc} 00 , & 10 , & 0000 , & 0010 , & 1000 , & 1010 , & \ldots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \ldots \end{array}$$

Here every group has one more new fix digit and thus the group totals will become half of the previous. So the full total is $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = 1$
This is a good point to pause and talk about infinite sums.
The ancient Achilles paradox was about this fact that the sum of infinite many numbers can be finite if they diminish. Today this is served on the silver platter of abstractions and so we seemingly grew above it. If I stand one meter from the wall and I go halfway, then half meter has been moved and half is still left. Then if I go a quarter meter then again a quarter is left. Going again half ahead means going an eighth meter and leaving an eighth meter from the wall. Doing this halfing infinitely I clearly approach the wall and so the one meter becomes the sum of the infinite many half plus quarter plus eighth and so on. The infinite decimals are an even more general method of combining numbers from infinite many smaller and smaller parts. This infinite sum vision is fully accepted in space. Time is a bit more difficult to be regarded this way. For example the idea that I can really approach the wall in the above manner would suggest a faster and faster approach or would suggest an approach that takes for ever. And yet it’s not true. Going with fix speed, the same infinite sum of space defines one in time too. The historical view of things happening this or that many times must be altered to accept infinity too. If it starts to rain now, then in the last minute, there were infinite many points in time when it wasn’t raining. With this view then the Achilles paradox is resolved. A faster object surpasses a slower, even though infinite many times before this surpassing it was behind. Infinite many times doesn’t mean always or forever. If a faster object starts from behind a slower then when it reaches the start of the slower, that slower will be again a bit ahead. Reaching this point, the slower again will be a bit ahead. And so on, we can locate infinite many points in space and so also in time, when the faster object was still behind. But that doesn’t mean to be behind forever, merely that these points approach the crucial surpassing point.
The Anti-Achilles paradox is the fact that the reversal of the Achilles is not true!
Smaller and smaller numbers don’t always add up to a finite, they can be infinite too!

To see this, we should first use non decreasing, rather same members.
Obviously $1 + 1 + 1 + 1 + \; . \; . \; . \; .$ gives infinity. Then we can decrease these members by replacing them with more and more new members. The second 1 could be replaced by two, the third by three and so on new members. The most obvious replacement would be two halves, three thirds and so on:

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

Then to make them different is easy! Keep the last ones in each group but increase all the earliers a tiny bit more and more, so that the first is still under the last in the previous group. Later we show much simpler example for such diminishing numbers that add up to infinity.
Diminishing numbers that still add up to infinite, could be called slowly diminishing.
These are the beginning chances that in spite of diminishing don't warrant a stopping.
An "argument" for this could go as follows. We go forward and group more and more chances to always get above a value say half. Then each group has a half chance in total, meaning a half chance for the "or" value of the group, that is at least one succeeding. In this infinite sequence of groups they all have more than half chance, so infinite many must succeed. But this means at least one in every group so also infinite many in the full sequence. This argument was faulty in so many details that it is hard to start where it was wrong. The chances of "or"-s are only bounded by the sums and equal only for exclusive events. Bounding was not enough here, we needed the "or" chances all to be above half so infinite occurrence could be claimed by the fix chance repetition and a hidden use of implication too. But the beginnings are not exclusive, they have to continue each other. And yet the grouping idea has some merit to it as we will see later in the exact proofs. Most amazingly we can have a second go and try to refine the same idea but it will be actually more detailed mistakes.
The basic idea is to avoid the need for exclusions by our old trick of calculating the "or"-s as complements of the "and"-s of the complements. So we again create groups but they will be "and"-s of the failings. This is good logically because having at least one true in an original group means the new group being false as all members failing.
But "and" is calculated by multiplication. So not only we lost the addition but instead of the original chances we have to use their complemented that is one minus values.
Amazingly, a classical theorem claims that infinite sum value of chances is exactly equivalent to zero total product of the complemented chances:

$$p_1 + p_2 + \; . \; . \; . \; = \infty \quad \Leftrightarrow \quad (1 - p_1)\;(1 - p_2)\; . \; . \; . = 0$$

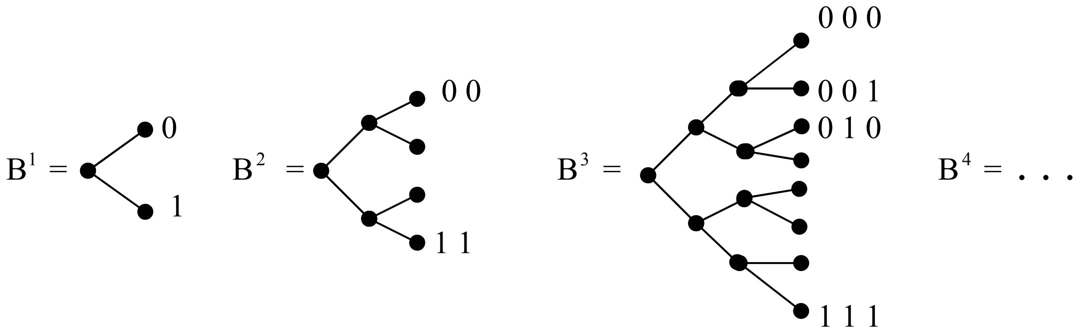
Thus we can form groups with arbitrary small fix product values, say all under half.
This means more than half chance of failure for the "and" of the negatives, that is more than half chance of success for at least one member.
The problem is that now, this product rule of "and"-s is only a lower bound in general and only equal for independent events. We needed equality but the beginnings are not independent! Most amazingly, after all these false arguments, a seemingly much easier but actually much deeper correct one shows what we really need:
Finite total of B implies the stopping of random sequences in B.
The first non deep, merely surprising fact is that if a sum is finite then the "end sums", that is the totals from later and later points, must approach zero. Indeed, the beginnings approach the finite sum, so the ends are the "left over" values that must become arbitrary small! In fact, this end sum zeroness can only happen if the total is finite. Indeed, if the sum is infinite then the left over end sums are all infinite too. So the "barbaric" totaling of all chances in a B beginning set has its more probabilistic meaning as the longer and longer beginnings having diminishing chances in total.

Next we can use the unconditional law that the sum bounds the “or” but now for infinite many members! Thus the chance of even just one occurrence after a point is bounded by the end sum from that point! So if these are diminishing then the chances of even one occurrences become arbitrary small too. Finally, we can regard infinite many occurrences as a single event that implies all of these later and later “or”-s.
So, we can use the law of implicative chances infinite many times! With the fix infinite occurrence as cause and with all possible later and later “or” occurrences as consequences. The cause can have no more chance than any of its consequences. So this cause, the infinite occurrence has chance not more than arbitrary small values. But only zero is not more than arbitrary small values, so the infinite occurrence has zero chance. Finally the crucial “deep” step is taking this zero chance as impossibility! This means exactly the claim that only finite many of the chances come true.
This gives the big justification for regarding the finite total chance of B as the necessity of finite occurrence of B elements in a random sequence. Or to put it negatively, an infinite occurrence from a finite totaled B must be a strangeness, only possible for mathematically created sequences.
To recap our results even clearer:
We’ll use B collections of beginnings as strangenesses!
A qualitative and quantitative finiteness is required for these B sets.
The qualitative finiteness means being effectively that is machine determined.
The quantitative finiteness of B is the above explored finite chance total.
Such both qualitatively and quantitatively finite B beginning sets are the forms of strangenesses. For an s sequence to be strange means a third counter infiniteness against the two finiteness of B, namely to have infinite many beginnings from B. This, we also call as “obeying” B. To only have finite many beginnings from a B, that is to stop having beginnings from B is thus disobeying or failing B.
So, only a strange s obeys a strange B, while the random sequences must all fail it.
The claim that our special B strangenesses can catch all strange sequences means:
If an s sequence fails all strange B then it is random.
Using only beginnings is a very strong step back from the heuristic P(s).
The first bonus is that the generally mysterious split of universality and strangeness is now fix. Obeying is strange and disobeying or stopping is the universal.
Subjectively this sounds good! To recognize some beginnings in an s sequence infinite many times already sounds like a strange thing.
But lets list all beginnings again:

0 , 1 , 00 , 01 , 10 , 11 , 000 , 001 , 010 , 011 , 100 , 101 , 110 , 111 , 0000 , . . .

This is finitely determined and yet every sequence will obey it. So we need a narrowing of this. We need the quantitative finiteness. We already revealed its form as the finiteness of the total chances but now we repeat the arguments precisely.
An other example of failure helps to start. Let B be all beginnings that end with a 0. In a random sequence, half of the beginnings in average should be like this. So obviously, must be infinite many. So, the random sequences obey B. Where did we go wrong? What made the randomness go through B infinitely? To see it, first lets formalize how we listed all beginnings above. We went by increasing lengths.
In general, B^n can denote the n long beginnings in a B collection.
That is, $B^n = \{b \ ; \ <b> = n\}$ where $<b>$ denotes the length of b.
This B^n is clearly finite, namely $B^n \le 2^n$. Indeed, every digit has two choices, so an n long beginning can be 2^n kind. We can see this in our above listing of all beginnings too. They went exactly in 2^n groups, that is 2 , 4 , 8 , . . .

Any B can also be listed in that order as: $B = B^1 \cup B^2 \cup B^3 \cup \ldots$
These B^n members could be listed in increasing binary order, as we did in our listing of the full set of beginnings. But, it's much better to visualize the B^n-s vertically, in fact, as trees:



We marked the 2^n branchings but, the end markings are only placed at the paths that correspond to beginnings belonging to B. As we see, these paths don't have to relate to each other in any manner. Indeed, B was arbitrary. The above mentioned 0 ending collection as B, broken into B^1, B^2, . . . will show a very simple rule: $B^1 = \{0\}$, $B^2 = \{0\,0, 1\,0\}$, $B^3 = \{0\,0\,0, 0\,1\,0, 1\,0\,0, 1\,1\,0\}$, and so on.
We have exactly half of the branches. This is obvious from the property. But, more importantly, in every B^n, half of the cases are the previous paths continuing with 0 and the other half are the non previous ones plus 0. This leads to the fact that for a sequence, we always have a new $\frac{1}{2}$ chance to belong to each B^n. An other way to say this is that for the sequences to belong to the B^n-s are independent, so their chances are exactly like flips of a coin. Then of course, just like at the coin flips, randomness has to find its way to infinite occurrence. In fact, we expect that a random sequence would belong to half of the B^n sets, just like half of the coins should be head or tail. But for us now, merely the infinity of occurrence is important.
Any windows that have all chances bigger than any fix ε, will allow randomness to pass through and obey. Does that mean, that diminishing chances always shut the door to randomness? Unfortunately not! Look at the following probabilities:

$$\frac{1}{2} \quad \underbrace{\frac{1}{4} \quad \frac{1}{4}}_{\frac{1}{2}} \quad \underbrace{\frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8} \quad \frac{1}{8}}_{\frac{1}{2}} \quad \underbrace{\frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16} \quad \frac{1}{16}}_{\frac{1}{2}} \quad \frac{1}{32} \quad \ldots$$

Every block has $\frac{1}{2}$ total chance. But chances don't add up, even for independent trials, only for exclusive ones. So the actual chances of "or" that is having at least one success in every block, is a little bit less than $\frac{1}{2}$. Still, lets entertain that this little error wouldn't matter. Then, indeed, every block having at least one success, would have to succeed infinitely, so we would have infinite many blocks, with at least one succeeding and thus, infinite many successes too.
Amazingly, we can salvage the basic idea if we increase the probabilities themselves a tiny bit. Not all of them. The last ones in each block can remain the $\frac{1}{2^n}$ values, and the rest can become all the reciprocals:

$$\frac{1}{2}\quad \underbrace{\frac{1}{3}\quad \frac{1}{4}}\quad \underbrace{\frac{1}{5}\quad \frac{1}{6}\quad \frac{1}{7}\quad \frac{1}{8}}\quad \underbrace{\frac{1}{9}\quad \frac{1}{10}\quad \frac{1}{11}\quad \frac{1}{12}\quad \frac{1}{13}\quad \frac{1}{14}\quad \frac{1}{15}\quad \frac{1}{16}}\quad \frac{1}{17}\quad \cdots$$

The blocks seem to be pointless now, but lets keep them anyway.
Lets also look at the complement $1 - p$ probabilities:

$$\frac{1}{2}\quad \underbrace{\frac{2}{3}\quad \frac{3}{4}}\quad \underbrace{\frac{4}{5}\quad \frac{5}{6}\quad \frac{6}{7}\quad \frac{7}{8}}\quad \underbrace{\frac{8}{9}\quad \frac{9}{10}\quad \frac{10}{11}\quad \frac{11}{12}\quad \frac{12}{13}\quad \frac{13}{14}\quad \frac{14}{15}\quad \frac{15}{16}}\quad \frac{16}{17}\quad \cdots$$

The product of every block now is the chance of "and" , that is total failing in the block, not occurring any of the trials. Lets calculate these:

$$\frac{1}{2}\quad \underbrace{\frac{2}{3}\bullet\frac{3}{4}}_{\frac{1}{2}}\quad \underbrace{\frac{4}{5}\bullet\frac{5}{6}\bullet\frac{6}{7}\bullet\frac{7}{8}}_{\frac{1}{2}}\quad \underbrace{\frac{8}{9}\bullet\frac{9}{10}\bullet\frac{10}{11}\bullet\frac{11}{12}\bullet\frac{12}{13}\bullet\frac{13}{14}\bullet\frac{14}{15}\bullet\frac{15}{16}}_{\frac{1}{2}}\quad \frac{16}{17}\quad \cdots$$

Thus, not only these must come through infinitely, but they have to fail infinitely too.
So, in infinite many blocks, we must have at least one success.
Thus indeed, the original sequence has to have infinite many successes too.
So in diminishing probability sequences of independent trials, we can create blocks of fix chances, and thus, randomness passes through with infinite occurrence.
Now the question is whether this can be done in any diminishing sequence or not. Not only, there is a crystal clear split where it can or can not, but that splitting condition will be simpler about the $p_1 , p_2 , p_3 , \ldots$ original chances, than about the $1 - p_1 , 1 - p_2 , 1 - p_3 , \ldots$ complements that we used.
To create a k block after n, we need that:

$$(1 - p_{n+1})\ (1 - p_{n+2}) \ldots (1 - p_{n+k}) \ \le\ \frac{1}{2}$$

Indeed, then not occurring any of the $p_{n+1} , p_{n+2} , \ldots , p_{n+k}$ has maximum $\frac{1}{2}$ chance, so at least one occurs with at least $\frac{1}{2}$ chance.
Thus indeed, the repeated occurrences in each block have at least as a chance as a coin flip and will infinitely happen.
Of course, instead of $\frac{1}{2}$ any ε would be enough, that is:

$$(1 - p_{n+1})\ (1 - p_{n+2}) \ldots (1 - p_{n+k}) \ \le\ 1 - \varepsilon \quad \text{is enough too.}$$

As we'll see, it doesn't make any difference in the end conditions either.
The full infinite product $(1 - p_1)\ (1 - p_2) \ldots = \Pi$ can be any value from 0 upto 1. But while 1 can only be with all $p_n = 0$, the 0 value can happen with seemingly any p values. Indeed, multiplication brings the:
$(1 - p_1)$, $(1 - p_1)(1 - p_2)$, $(1 - p_1)(1 - p_2)(1 - p_3)$, $\ldots$ values, lower and lower, so going down to 0 is quite natural.

For example, at the $p_1 , p_2 , \ldots = \frac{1}{2} , \frac{1}{3} , \frac{1}{4} , \ldots$ reciprocals' complements:

$$(1-p_1)\ (1-p_2)\ \ldots\ (1-p_n) = \frac{1}{2}\ \frac{2}{3}\ \ldots\ \frac{n}{n+1} = \frac{1}{n+1} \rightarrow 0$$

The more natural doubt is whether this limit can be $\Pi \neq 0$.
The easiest example is the reciprocals of the squares:

$p_1 , p_2 , \ldots = \frac{1}{4} , \frac{1}{9} , \frac{1}{16} , \ldots , \frac{1}{n^2} , \ldots$ Indeed:

$$(1-p_1)\ (1-p_2)\ \ldots\ (1-p_n) = (1-\frac{1}{2^2})\ (1-\frac{1}{3^2})\ \ldots\ (1-\frac{1}{n^2}) =$$

$$\frac{2^2-1}{2^2}\quad \frac{3^2-1}{3^2}\ \ldots\ \frac{n^2-1}{n^2} =$$

$$\frac{(2-1)(2+1)}{2\bullet 2}\quad \frac{(3-1)(3+1)}{3\bullet 3}\ \ldots\ \frac{(n-1)(n+1)}{n\bullet n} = \frac{(2-1)\,(n+1)}{2\,n} = \frac{n+1}{2\,n} \rightarrow \frac{1}{2}$$

Now that we see that Π indeed can be anything, comes the surprise:
The limit of the "end" products, can only be 0 or 1:

$$\lim\,(1-p_{n+1})\,(1-p_{n+2})\ \ldots = \begin{cases} 0 \\ 1 \end{cases}$$

Indeed, this L limit is $\lim \dfrac{\Pi}{(1-p_1)\ \ldots\ (1-p_n)}$.

Now if Π was 0, then every member itself is 0, so the limit too.
If $\Pi \neq 0$, then the members are not 0 but the $(1-p_1)\ \ldots\ (1-p_n)$ "denominator" values approach Π, so the "fraction" approaches 1.
This shows that the choice of $\frac{1}{2}$ for blocks is immaterial.
To go under any ε from any $(1-p_{n+1})$ can only be done if the end product $\rightarrow$ 0.
So the end product being 0 or 1 gives the sharp distinction.
This also means the $\Pi = 0$ or $\Pi \neq 0$ sharp distinction for the full product.
But we promised something for the p values.
Indeed, it's tedious, but quite elementary to show (see Occurrence Theorems) that:

$$(1-p_1)\ (1-p_2)\ \ldots = 0 \quad \Leftrightarrow \quad p_1 + p_2 + \ldots = \infty .$$

The reciprocals are a perfect example for this by the way:
We already showed that the left side is true:
$(1-p_1)\ (1-p_2)\ \ldots = \frac{1}{2}\ \frac{2}{3}\ \frac{3}{4}\ \ldots = 0$

The right side can be seen easily by:

$$\begin{array}{cccccccccccc}
\frac{1}{2} + & \frac{1}{3} + & \frac{1}{4} + & \frac{1}{5} + & \frac{1}{6} + & \frac{1}{7} + & \frac{1}{8} + & \frac{1}{9} + & \frac{1}{10} + & \frac{1}{11} + & \frac{1}{12} + & \ldots\ldots\ldots\ldots = \infty \\
 & \vee & & \vee & \vee & \vee & & \vee & \vee & \vee & \vee & \\
 & \frac{1}{4} & & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \\
 & \multicolumn{2}{c}{\underbrace{\hspace{4em}}_{\frac{1}{2}}} & \multicolumn{4}{c}{\underbrace{\hspace{10em}}_{\frac{1}{2}}} & \multicolumn{5}{c}{\underbrace{\hspace{16em}}_{\frac{1}{2}}}
\end{array}$$

So, we obtained the simple result that in a B too, some kind of tricky grouping of beginnings can only give infinite many fix chance windows if the total chances in B are infinite. To be precise:

$< b >$ denotes the length of b. For example: $< 0\ 1\ 0 > \ = \ 3$.

$|\, b\, |$ denotes the chance of b, which is $|\, b\, | \ = \ \dfrac{1}{2^{<b>}}$.

For example, $|\, 0\ 1\ 0\, | \ = \ \dfrac{1}{2^3} = \dfrac{1}{8}$. And indeed, $\dfrac{1}{2} \bullet \dfrac{1}{2} \bullet \dfrac{1}{2}$ is the chance.

By the way, if the trials are not coin flips, rather trials of a p probability, then:

$|\, 0\ 1\ 0\, | \ = \ (1-p)\ p\ (1-p)$. For example, for a dice: $|\, 0\ 1\ 0\, | \ = \ \dfrac{5}{6} \bullet \dfrac{1}{6} \bullet \dfrac{5}{6}$.

The total of the chances in a B is $< B > \ = \ \sum\limits_{b \in B} |\, b\, |$.

This itself is not a chance or probability of B, because many b in B are not exclusive. And indeed, $< B >$ is usually much bigger than 1.

As it turns out, the crucial difference is whether $< B >$ is finite or infinite.

If $< B >$ is infinite, then it might be possible that fix chance gates allow randomness through. We didn't pursue these constructions exactly, but we showed that:

$< B > \ = \ \text{finite}$ definitely excludes such.

Amazingly, $< B > \ = \ \text{finite}$, not only excludes these not quite exact grouping ideas, but exactly guarantees the opposite. Only finite many from B can be in a random s.

Most amazingly, the "proof" of this is simpler than all the aboves. We don't need complement chances and products All we have to do is apply our basic rules of chances to infinite many members and then regard 0 chance as impossibility.

The most basic rule is that if a property implies a consequence, then its chance can't be more than the chance of the consequence. The other is that the chance of an "or" can't be more than the total of the members. These two rules can form a chain if the consequence of a property is the "or". Indeed then:

$$p\,(\text{ property of } s\ ) \ \leq \ p\,(\ \mathbf{1} \text{ or } \mathbf{2} \text{ or } \ldots\ ) \ \leq \ p\,(\mathbf{1}) + p\,(\mathbf{2}) + \ldots$$

Now if $< B > \ = \ p_1 + p_2 + \ldots = \ \Sigma$ then,

$$\lim\,[\ p_{n+1} + p_{n+2} + \ldots\ ] = \lim\,[\ \Sigma \ - \ (p_1 + p_2 + \ldots + \ p_n\,)\ ] \ = \ 0$$

Indeed, the $p_1 + p_2 + \ldots + p_n$ beginning sums approach Σ itself.

So the end-sums are arbitrary small. Now we apply the above chain as follows:

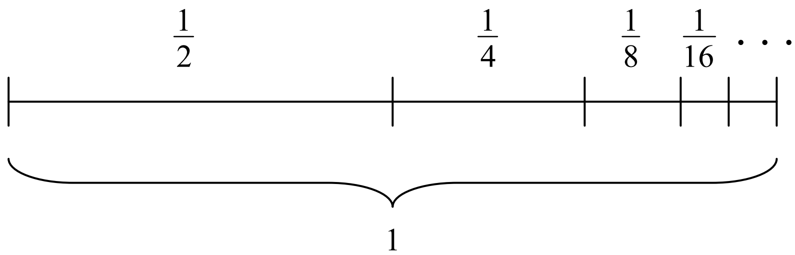
The mysterious "property" should be having infinite many beginnings from B.

And the **1** , **2** , . . . cases are p_{n+1} , p_{n+2} , . . . each coming true.
Indeed, then our property, that is infinite many from B implies the "or" of these cases, for any n. So actually we have infinite many applications of the chain, by regarding the different ends in our sum. They become arbitrary small, so:
p (s has infinite many beginnings from B) = 0. So for a random s, it is impossible.
Thus, we found the promised narrowing of B-s as: $< B >$ = finite.
This is the quantitative finiteness of B and the perfect condition of strangeness is:
B is finitely determined (quality) , $< B >$ = finite (quantity) , s obeys B (infinity)
An obvious consequence of $< B > = p_1 + p_2 + \ldots$ = finite is that $p_n \rightarrow 0$.
This means that for any $\varepsilon > 0$, there is an N, after which, that is for $n > N$, $p_n \leq \varepsilon$.
Indeed, the opposite of this would mean having an ε value so that p_n don't stay under ε from any N. Then, we would have infinite many p_n values above ε . And then of course, the total $< B >$ would have to be infinite.
But the reverse is not true. $p_n \rightarrow 0$ does not imply $< B >$ = finite.
Indeed, we saw that $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \infty$.
The $p_n \rightarrow 0$ sequences can be called as diminishing. Where the total is ∞, should be called slowly diminishing, while the ones with finite total, as fast diminishing.
Two examples of this are:

$$\frac{1}{2} , \frac{1}{4} , \frac{1}{8} , \ldots , \frac{1}{2^n} , \ldots \quad \text{and} \quad \frac{1}{4} , \frac{1}{9} , \frac{1}{16} , \ldots , \frac{1}{n^2} , \ldots$$

The calculation of the exact finite sums of fast diminishing sequences is very strange.

Sometimes it's trivial, for example, the first $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$ means:



This is so, because the leftover from 1 after every beginning sum, is exactly the last member. Indeed, it's true for the start and inherits to new sums. Or directly:

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

The other example we mentioned on the other hand, has a strange sum:

$$\frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots = \frac{\pi^2}{6} - 1$$ which took years for Euler to establish.

The complement product was easy as we showed to be $\frac{1}{2}$, so the finite sum followed from the general law of sums and complement products. But not the exact value.
Now I want to show how useful the B strangeness is, especially if we regard B not as an effectively given set, rather some beginning property.

The $< B > =$ finite condition is simple, but if B is given as an actual beginning feature, rather than a collection, then an easier way to establish $< B > =$ finite would be to examine: $< B^1 >, < B^2 >, \ldots$

First of all, since the B^n sets are disjoint, that is have no common elements, thus the total $< B >$ value is exactly the sum of the $< B^n >$-s:

$$< B > = < B^1 > + < B^2 > + \ldots$$

Then with $< B > =$ finite, these must diminish too, that is approach 0.

In fact, with our above nomenclature, $< B > =$ finite , that is our kind of condition of strangeness, means that $< B^n >$ has to diminish fast.

This sounds like "fancy talking" because in the end it means the same, having finite total, that is $< B > =$ finite. But the fast diminishing of $< B^n >$ is an "intuitive base", that justifies the finite occurrence in random sequences much more than the formal "proof" we gave above with the impossibility of the 0 chance property.

Next I want to attack the old myth, that randomness is all about the "Law of Large Numbers". By the sound of it, this law should be about the mentioned fact that with big enough number of trials, everything is possible. We can throw even a million heads in a sequence, if we try and try new million long throw sequences. This of course would take a very, very long time. This intuition I fully accept as a priori, hail Kant and god bless Plato. But this intuition doesn't have a special name. I will call it the "Law of Big Numbers". Instead of this, the Law of Large Numbers refers to the "intuition" that in long enough trial sequences, the heads and tails even out, that is appear about the same many times. Or as they say, their frequencies are about the same. Of course, the previous intuition, the Law of Big Numbers tells, that they don't always even out, they can be arbitrarily uneven too. So what we really mean is merely that these uneven situations are much rarer than the even ones. So if we group all possible outcomes of a million long coin flips, according to the head tail differences, then the small difference groups, say from 0 to 100 difference, will have much more cases than the big difference say from 1,000,000 to 1,000,000 – 100 = 99,900.

We can already see this at small numbers.

For example, in the ten coin flip sequences, the 0 and 1 differenced ones have

$\binom{10}{5} + \binom{10}{4}$ cases, while the 10 and 9 differenced ones have $\binom{10}{10} + \binom{10}{9}$.

Here, the $\binom{\ }{\ }$ denotes how many choices of the lower can be from the upper many.

Clearly, $\binom{10}{10} = 1$ and $\binom{10}{9} = 10$, so only 11 in total, while $\binom{10}{5}$ and $\binom{10}{4}$ are each, much, much more. So the evened out cases are much more than the unevened.

At a dice of course, we go for the 6, which is not supposed to even out with the non 6 cases at all. In these, not half half chance situations, the evening out means that the $\frac{1}{6}$ proportional 1 occurrences, should be the majority of the cases.

Observe that all these distributions of cases, can be examined totally precisely by merely going through the possible cases mathematically.

We don't have to do experiments. Or if we do, we would get the same results!

The next step from this intuition of evenness in finite cases, is to go to infinity and claim that any uneven case collection tends to 0. This, is already a false claim.

Indeed, taking more and more trials, the uneven cases can and usually should increase and only their proportion to the total cases should diminish. But this is hard to calculate , so instead, we should grasp something that doesn't depend on the huge possible outcomes from n trials, rather only on n. Well, it seems easy.
We simply count the successes that is 1-s, and this k is the frequency.

Then, $\frac{k}{n}$ is the relative frequency or proportion of success. We already used the intuition, that this $\frac{k}{n}$ is about the probability p, like $\frac{1}{6}$ for a dice.

This "about"-ness was not pressed at that time, but if we try to exactify it, then it goes towards two directions. One, is about in the outcomes of a fix n many trials. And one is about how it changes, if n grows.
So now we can combine these two and claim that it also incorporates the previous evenness principle. We can do all this, but it doesn't make it plausible. We simply force feed people to believe that all this is good and dandy and we capture the essence of infinite tendencies. It's pure religion in math. But worst of it, is that it is preached in "latin". Indeed, after replacing everything with $\frac{k}{n}$ being close to the probabilities, they formalize this claim as $\frac{k}{n} \to p$ or $\lim_{n\to\infty} \frac{k}{n} = p$.

Then of course, the whole "church" of calculus comes in. The general $\frac{k}{n} \to p$ limits are easily reduced to $(\frac{k}{n} - p) \to 0$. So the essence of all this hidden "scripture" is what really diminishing means. It will turn out that it is not correctly named diminishing at all. Indeed, diminishing sounds like something happening infinitely often, while it actually means something happening only finitely.
But first lets start with real intuitions used to broaden our minds and demolish false ones. As we said, in finites, everything is possible, The Law Of Big Numbers is an intuitive fact. Lets combine this with a new intuition, namely that after a beginning, the next segment is totally new and has nothing to do with what happened in the beginning. This means that if in an s sequence, we mark <u>all</u> the $b_1, b_2, \ldots$ beginnings that satisfy some B beginning property, then looking the segments after $b_1, b_2, \ldots$ we must have all segment variants. The marking of <u>all</u> B beginnings was very important. Indeed, if we just choose from B kind of beginnings, then we can cheat by choosing ones that have particular segments after them. But if we choose all of them, then since a B property can't influence what comes after a B kind of b, thus in a random s too, the full variety of B beginnings must give the full variety of continuations. The only requirement is to have infinite many B beginnings.
A simplest application of this logic already destroys a common naïve mistake in our intuitions. Namely, we feel that after a lot of heads in a row, the tail should be more expectable. It is a misuse of the correct intuition about the rarity of full heads in a row. They are rare among say the hundred flip sequences. But when go to the hundred and one, then they are merely beginnings, and each can be followed equally with head or tail. Now, the intellectual destruction of false intuition pathways, is not as efficient as the naïve formings of new ones. So just because you are a mathematician, you could still have a creepy feeling, to rather wait a few spins of the roulette with reds before you place your bet on the black. It's best if you don't go to the casino at all, and instead just think about the chances.
Instead of a single outcome after the B beginnings, we can look at fix m segments as windows. Then, in these windows all possible m outcome segments must appear.

Now we inject a new third intuition to use as B.
In the coin flips, we not only expect that the number of heads and tails are about equal, but also, that the role of head or tail is immaterial.
So, same many excess 1 or excess 0 cases should be. Moreover, in a random s, then this excess should also alternate and as a consequence, we must have infinite many exact equalizations.
Now comes the application that will tell something amazing about the excesses in size.
Let our B property be that b is an equalization. So this means that $< b > =$ even and b has equal 0-s and 1-s. After the infinite many such b-s in a random s, lets place the m windows. Unfortunately, we could have new equalization in a window, so the windows would overlap. So lets go from the start, and always regard the first new windows and ignore the ones that would start inside. We end up with infinite many m windows that don't interfere and each follows an equalized b beginning.
The beginnings are special but since we took all special ones, except for the minor overlapping deletions, thus we could say that these positions are random. Or to put it an other way, since we took all these equalized beginnings, thus, there was no choice involved, so even accidental cheating by selecting special equalization, with considering what comes after, is impossible. So, the windows must display all possible outcomes, including full 1-s or 0-s. In fact, infinite many times.
So since m can be arbitrary big, we must have arbitrary big excesses of 1-s over 0-s or vice versa. So to say that the 1-s and 0-s equalize in the long run is totally false.
Exactly the opposite is true. They oscillate bigger and bigger.
Since the relative frequencies of both the 1-s and 0-s approach $\frac{1}{2}$, this also means that the relative oscillations, that is these increasing oscillations divided by the n, where they happen, must approach 0. So why should this be obvious? It isn't.
It is a new, even deeper consequence of randomness. First, we have to look closer, what an approach to 0 really means.
$p_1, p_2, \ldots \rightarrow 0$ or $\lim p_n = 0$ means that for every $\varepsilon > 0$ value, there is an N number that after N, that is for $n > N$, all p_n values are under ε.
So actually, for every ε, we have a different $B(\varepsilon)$ beginning property, that contains those b beginnings, in which the relative oscillation is above ε.
These are the "strange" more than ε relative oscillating beginnings.
In the B^n, we can check whether these are indeed rare. And they will be.
Not because of intuitions, but because of simple counting of cases.
So, $< B^n(\varepsilon) > \rightarrow 0$ fast. Which means that $< B(\varepsilon) > =$ finite, so $B(\varepsilon)$ is indeed a strangeness. But then, according to the rule of randomness, the strangeness must stop in a random s. It is merely happening a few times accidentally, it can't go on.
An s can't obey $B(\varepsilon)$, it must fail $B(\varepsilon)$. So after a last N long beginning of s, for all $n > N$ long b beginnings, $B(\varepsilon)$ will be false.
So the relative oscillations of B remain under ε.
The finite versus infinite is the true face of randomness.
The Law of Large Numbers, is only a particular application of the big intuitive principle, that strangenesses must stop in any random sequence.
I was in high school when I realized this new principle and the necessary shift from P (s) to B beginning properties. Instead of the $< B > =$ finite condition, I was obsessed with the $< B^n > \rightarrow 0$ fast version. As we go further, it will become clear that the ultimate question is the original $p_1 + p_2 + \ldots =$ finite versus infinite.
Randomness is nothing more than a failing attempt to exactly solve this original problem. The more we depart from the head on problem, the more little details we get

about particular cases of the finite or infinite solutions disguised as results about randomness. Unfortunately, that's the only road we have. But at least seeing the big picture, we won't be deceived. In high school, I was deceived, because I thought that, $< B^n > \to 0$ fast is the absolute concept of randomness. At this same time, many mathematicians were deceived on a much higher level. They never even regarded the single B idea of strangeness, instead followed the old idea of the Law of Large Numbers. The origin of this pursuit goes back to before the war. Before the idea of effectivity or machines even emerged.

The seemingly heuristic idea of Von Mises was to require $\frac{k}{n} \to p$ not only for the total s sequence, but for any s_0 subsequence. This in set theoretical sense, of course is clearly impossible, because s_0 could be all the 1-s or 0-s. The s_0 could also be given by its positions, that is as a sequence of natural numbers. It could be called an observational sequence. An observational sequence like $2 , 3 , 5 , 6 , 1 , 10 , \ldots$ would give for every s binary a new one, by simply regarding the second, third, fifth and so on digits in s. If the observational sequence doesn't repeat values, then this new binary is indeed a subset. The observational sequence can be itself a random or even ruled one. Luckily, it doesn't make any difference. The observed subsequence of an s random should be random in both cases. Indeed, if we pick by rule, then the random s shouldn't care about this rule, so we should get random s_0 subsequence. If we pick randomly, then even more.

So it sounds like we rose above random or non random, to grasp something absolute. But this is a delusion. Being ruled doesn't have to be absolute rule, rather, dependant on s itself. Then of course, we can select the places of all the 1-s or 0-s, so clearly we wouldn't get a random sequence. To exclude the dependence on s at all, is not viable either. Indeed, if a sequence has the weirdness of having 1 0 1 always after ten 0-s, then this can not be detected by a fix observational sequence, only the one watching for ten 0-s and then picking the next three digits. We could say that this is okay, because here we determined observational places from the beginning before it.

All this was pursued pro and con and finally, Church even introduced effective collection of beginnings to give observational places after them.

The crucial break away from this direction, used effectivity of the beginnings also, but in a totally different way. Only ten years after this best definition of Martin Löf came the surprisingly simple use of "obeying" as infinite occurrence by Solovay. He still didn't use a single B and only decades later did Rod Downey simplify it for such.

So it's like Playfair's beautiful axiom for parallelity, that we use today, and which replaced the ugly Euclidean version only after two thousand years. Here, it only took forty years to find the simple form of which I had the vision in high school. But behind that naïve vision, lies a very complicated failure of the whole search for absolute randomness. To expose this contradiction, is very hard. Not hard, because of its complicatedness, rather because of our stupidity to stay in the blind spots of Formalism. So next to the Well Ordering Theorem, this classical definition of Martin Löf randomness, is the most misrepresented, over simplified and abused concept of mathematics. Both of them meant personal coincidences for me. I met Paul Cohen intentionally when I was very young and the spookiness was something that happened due to my poor English. But just a year ago, I emailed Rod Downey a few times with questions, without even knowing that he was the one who simplified the Solovay definition to the form that I envisioned in high school.

To follow your vision and not bother about that has been achieved is a good strategy to become a real mathematician but eventually you will have to go out from your cave!

Especially, because the messages we receive from the other side, are very fragmented and unbalanced. We can see details of a far future and yet be blind to basic past. Still, this is the only way to get vision at all. This is so because the opposite direction of keeping to be well informed is placing us into the mercy of Formalism!
The logic of derivations completely ignores the unsuccessful paths, yet only these give the real meanings. Gauss was a crucial figure in this new menace but his idol Newton was already the origin of new Formalism. So in fact even he, merely followed an older road. As we dig back, we realize that Euclid struggling to establish the relation of the three forms of parallelity, finally forgot to explain the three forms.
To a total outsider it's quite plausible that lines that are:

1. Having the same fix distance between them, or
2. Having same angles to any third lines that crosses them, or finally
3. Never crossing each other,

mean the same three things and so we can use any of these as definition of parallelity.
The real business starts if we want to derive the sameness of these three.
The healthy formalism is to simplify and organize the results that our wonderings yield. But this healthy formalism slips into a devilish trap. We don't want to talk about the unsuccessful detours, the very road we traveled. Honesty still dictates that we reveal at least some depth of the original problems. Euclid was honest but still failed in two sense. If we could show him today the presently accepted Playfair axiom, it would yield amazing psychological insights. This new axiom says that to a given line we can only find one non crossing through any points outside the given line.
Seemingly, this concerns only the last third feature of parallels, non crossing. But in a ingenious way it actually encapsulates the crucial assumption that settles the equivalence of the three meanings. It's not that Euclid was not aware of this truth and subconsciously was also aware of that it implies his more complicated axiom. The reason it didn't spring to a conscious alternative instead of his, was simply the intention to reveal the problems. But we can not serve two lords at the same time!
Simplifications to more abstract forms must be allowed to evolve on their own course and not compromised by visual simplicities! The two simplicities must be combined in a totally separate endeavor as didactical exploration. Very narrow history of such endeavors does exist but it is intentionally limited and blocked. Namely, only unavoidably positive aspects of such didactical routes are allowed. The personal weaknesses of the geniuses is only one aspect. There is a teleological deeper reason, a failing of mankind as such. The present age is the full blown battlefield of this failing.
The duality of Academia and Media is the concrete form of total deception.
In the Academia Formalism rules, in the Media, patronizing oversimplification!
Society is an independent entity, with its own goals. The social morons, the slaves to social success are already denying the existence of such "oversimplified" entity. So the old saying that the devil's greatest trick is to make its existence ridiculous is true.
Of course there is a lot of modern talk about the devil. But always psychological evil is regarded. Bad criminals as the most primitive cartoon version or bad cops, bad judges, bad teachers, bad politicians in the social dramas. The system itself being bad and evil is rarely expressed and even if so, the consequences are softened.
Blocking understanding from the individuals is the most crucial evil of society!
The devil itself was only recognized in the modern times by Timothy Leary in his concept of the "Establishment". This non political entity can not be defined abstractly but felt at once in your stomach if you take LSD. The search for happiness, the politics of ecstasy jumps over the politics of Formalism and is an abstraction itself. The fact that common sense is still in us, is forgotten. So the new media, computers,

the internet was hailed by Leary as some salvation. That was the total betrayal of everything. A final sick twist, sinking to the role of a clown.
But we are lucky now because we deal with math and inside here, even the devil is more concrete! So every didactical fact is unavoidable truth and the Formalists can merely say “so what?”. The Academia is thus simply pretentious, ignorant to society outside, ignorant to teaching and learning too. This modern Formalism thus indeed can be seen at Newton already. Came from the bottom and wanted to be at the top.
Even his pretentious humbleness, hid a crucial denial. He said that if he could see farther, then it was because he sat on the shoulders of giants. The fact that what he saw was true not merely by verifiability but by understanding, is missing! And indeed, the crucial success of his vision became not understanding how forces cause accelerations, but the fact that this allowed the derivation of the Kepler laws. The insiders opinion was the only important point. And indeed, this is what put him to the top. The darkness of Jekyll and Hyde ruled his whole life. It was embarrassing that his niece helped the first entry into higher circles and it was a life long hidden goal to make gold, to become God in an even more absolute sense than social belonging. This sickness of the greatest genius of all time is actually a symptom of the social devil but to see this is very hard. We who look for the bigger beast than man, can be very objective because we can jump over these dirty details and look at the failings of understandings. These are pure and simple just like the abstract laws themselves.
Newton was reluctant to even publish his physics. But what he did publish is amazing! Overcomplicated details of directions that are dead today. And yet not one single didactical intention. The famous picture of the cannon on Mount Everest launching a “satellite”, was only included in later editions. This became now an obligatory didactical vision. Only allowed by the devil because the much deeper problems of vectors in calculus leaves it isolated anyway, so can not induce understanding in its own. The problem of energy as a metaconcept is completely suppressed from the origin of mechanics even today. The overcomplicated arguments of the Principia became “polished” to the oversimplified set of the Newton Laws.
The big admirer, Gauss created a new rigor in derivations without penetrating this rigor itself as Logic. The denial of visually convincing arguments, meant a general denial of visual understanding. A straight out lie, because he himself clearly had visual meanings that he kept to himself and replaced by exact abstract steps. No wonder they called him the “fox” because it also erases its footprints by its tail. The most famous of his results is the final proof of the Fundamental Theorem Of Algebra.
Among the complex numbers every polynom has root. In fact every polynom $P(z)$ will pick up all possible values as we try all possible values for z. The reason for this is simple! Each $a_k z^k$ member in $P(z)$ is merely spiraling outward k many times faster than z on the complex plane. Now, the different members may work against each other, but the biggest member will dominate all the others for large enough z, meaning the distance from the origin. Thus, $P(z)$ will spiral out to infinity eventually sweeping through everything. There is a fogginess in this because the possible z values are points of a plane, so in what order should we use them? But this can be made more exact even visually as follows. Suppose we want to show that the v arbitrary value is taken up by $P(z)$. We choose any other s starting point and envision all possible circles around this s. If $P(s)$ is v then we are finished, v is taken up as value. If not, then also there is a tiny circle around s so that its image by P is a tiny loop “around” $P(s)$. The tinyness is only required so that v is outside the loop. Now increase the tiny circle to bigger and bigger. For large enough, as we explained, the loop image is large too, so we can guarantee that the v value will be inside. But a loop continually increased can not become from one having v outside to

one having it inside, without crossing over v that is taking up the v value itself at a stage.
Seeing the topology of complex numbers was the crucial new edge that Gauss had over the previous purely algebraic meanings of mysterious real and imaginary combinations allowing determined values. He saw that the dual, one real and one imaginary sums will cover all possibilities. In fact, he believed that these duals or complex numbers are a new physical reality. He wanted to become the new Newton.
And he was right! Complex numbers are the foundation of modern physics, but he was too early to discover this physics yet.
So it's the same story over and over again!
Newton locking himself into his lab to make gold, Gauss dreaming of a new physics, Einstein spending decades on the new unified theory, Cohen trying to prove the independence of the Riemann Hypothesis. God gave these people precious insights to prove themselves once and hoped that they will turn to the bigger picture and recognize God in helping Man to stand up against the devil. Instead, they believed that what they achieved once can be repeated by will.
The failures of geniuses is the real history of Man's failing! And history will be rewritten! The saddest part is that these failing geniuses became bitter human beings too! Their twisted mind justified their isolation and they didn't realize how much opportunity was given to them to do something totally different and positive. This goes for all socially successful people! Fame, power and money has only one purpose! To turn against society where it is seemingly the least destructive yet where it "hurts" society the most. Cut through the Formalism, reveal the social lies, let the true details of knowledge be accessible to common sense. But knowledge is the most precious privilege and people want to be entertained. So the perfect duality of lies is sustained.
Feelings, bliss can not escape this contradiction. Thought is not superior to feeling but through the controls of thoughts society can control feelings. The economic and technological overwhelmingness of society can only be chiseled away by understanding. Art is trap! Artists should demand didactical science! Instead they form their own lies about science!
I have to return now to randomness because honesty requires to tell some details that I only crystallized later. Not through my naïve beliefs that came from the other side and neither by the understood latest status quo. The lingering or reemergence of certain ideas is nowhere as puzzling as in this field.
The Law Of Big Numbers as I called it, means that every finite combination must come about if we try combinations again and again. Most importantly, this is intuitive, it's an a priory of Kant. Throwing down hundred dices again and again we'll get all combinations eventually. Throwing dices infinite many at once or in a sequence in time has an additional minor element namely that it is also a set of repeated blocks. Once we get over this, the previous law tells that all possible blocks must appear. Regarding the twenty six letters of the alphabet plus the basic symbols like period comma and so on with an added space bar, that is using a typewriter randomly, we also should get all possible combinations, that is texts. More than hundred years ago Borel formulated this by claiming that if a monkey is using a typewriter for ever randomly then eventually he will write down the Bible in full too, merely as an accidental combination of hits. The shockingness of this might question our beliefs in a priories, but I don't think so! I think a priories always have this testing field that they must overcome. So we have these secondary paradoxes that must be resolved! Borel's monkey is combining the natural Law Of Big Numbers with the windows in a single sequence and thus becomes paradoxical due to the "extreme" consequence contradicting some practical experiences about monkeys. Amazingly, the time that is

sequence or window component is not necessary to create such paradox. Without infinite trials, the singular application of the Law Of Big Numbers can already be paradoxical. In fact, my son Daniel formulated this paradox when I explained him how real PC-s relate to theoretical machines. I missed it and merely emphasized how the infinity of the memory is the real step to abstractness while the operative systems of even the earliest computers were already universal. As a proof to show this point I simply argued that the behavior of a PC is determined by the totality of its bit configuration. Now since this is finite for a fix PC, thus after a while a fix PC would repeat itself and thus could not really diverge and thus miss the most crucial distinction of theoretical calculations. So I did use time as a factor to show this point and he said "It's like Borel's monkey". I said no, not at all, there we only regard one configuration, here the point is to get wider and wider. And then he finally was able to convey to me that to him it is already surprising that the "life of a PC " is merely a set of configurations. And then I realized that it is a paradox to me too that I simply jumped over. Every film and music you download to your PC could be produced by yourself if you could establish the exact bit configuration of those contents.
An earlier realization of randomness in actual nature was statistical mechanics.
The atoms and molecules in a gas like the air around us, are actually tiny billiard balls and the sameness everywhere is merely the result of random bouncings. This doesn't forbid that more particles could go to a certain place than usual but in that case being more there, will cause more likely reflection of new incoming ones. So basically everything is possible but the more uneven distributions are less likely. This means though that waiting long enough, even the unlikely distributions must come about. I don't know which would take longer time, for a monkey to write down the Bible or the air going from one room to the other accidentally. A much more interesting application of this molecular randomness is the melting and dissolution of a sugar cube in water. We never see a sugary water to go reverse and form a sugar cube. If all processes of the universe go towards the more probable random distributions, then all chunkiness or concentrations should come from the start of the universe and eventually dissolve. But we do create sugar cubes, even plants grow and use the water and air around them to build chunkiness. Can life and then intelligence even more, defy the probabilities? From the plants' example we soon realize that the suns energy is involved. But energy is not enough in itself! After all, a gas is containing energy already in its molecular motions. The fact that heat goes only from higher temperature to lower, shows that even a huge reservoir of molecular energy like the ocean will steal rather from the smaller energy containers than give up its own. And indeed just a minute temperature drop of our oceans would give enough energy for mankind forever.
Elemental refusals to this rule of randomness are obvious mathematically as one directional mechanical devices. Simplest is a trap door. Placed on a billiard table it would direct all balls to one side very soon. So why can't we build a molecular trap door? Such and similar devices or "demons" were envisioned by Maxwell way before Borel's monkey. So what makes these impossible? Nature! That's where math and physics departs. Way before Quantum mechanics and Relativity, we knew this frame law that forbids the tempering with the randomness of "real" billiard balls. Today anybody can design Maxwell demons and then the newer physical laws will interfere and defy his design. Unless we were wrong and we can steal energy from a random particle reservoir.
All these amazing thoughts show that not only the relationship of math and physics in general is a mystery but that randomness is at the heart of all. So it's no wonder that all the smart mathematicians were secretly struggling with these problems.

The “secretly” is appropriate because there was no apparent success, or official line!
In a sense there was no reason to even care about randomness. The calculations of combined probabilities was a success and it made real connection with already existing fields. So instead of separating randomness or strangeness, detailed combinations became provably 0 or 1 in chance. This was left to be interpreted as impossibility or certainty and could be speculated further towards randomness.
Of course in everybody’s back of mind, the actual randomness still remained.
Kolmogorov is the best example. He axiomatized probability theory and his line slowly but surely took over the speculations of Von Mises.
So Kolmogorov didn’t even claim the Law Of Large Numbers as fact merely as tendency. He didn’t say that the relative frequencies must approach the probabilities, rather that the chance of them approaching it is 1. This of course only makes sense if we build new chance calculations for whole trial sequences as events. And indeed he did and it opened up new deep or rather hard problems. Stronger and stronger versions of these approximative Laws Of Large Numbers were created.
An earliest of his such tendecysation of old facts is the zero one law. This is a generalization of Borel’s monkey. Again, Kolmogorov doesn’t claim that the monkey will actually write down the Bible, rather that it has 1 chance that it will. The advantage of this more careful claim is that it can be generalized. So what we loose in physical concreteness we gain it in actual scope. It turns out that the main reason the arbitrary text must appear in a random sequence is the fact that chance of such appearance is beginning independent. Every beginning independent or end property must have zero or one chance. But wait a minute, isn’t this a repetition of the mentioned heuristic $P(s)$ end properties? That failed because we don’t know what properties are and it can not be used with sets because random sets could be used to define sets. In reality we still know properties and can decide which of P or $\neg P$ is the strangeness. Indeed now with Kolmogorov’s zero one law it is the zero chance.
But Kolmogorov’s law avoids the philosophical trap. It avoids the whole property or set distinction because it talks about the independence of the chances already.
Of course, teaching these probability laws without the backdoor struggles of randomness is insanity. And indeed, all probability text books are insane!
Kolmogorov himself stayed with randomness continually! Martin Löf was his student when he found the new application of Effectivity avoiding the whole Von Mises direction and rather regard directly machine determined diminishing predictions as strangenesses. Most interestingly Kolmogorov believed in an alternative approach, using machines to define compressibility and regard randomness as non compressibility. The big western champion of this direction became Gregory Chaitin and we come to this line later.

3. Predictions

A seemingly much stronger new "obeying" can be defined.
Instead of "dipping" into a single pool of B beginnings infinite many times, the s sequence can contain at least one beginning from every B_1 , B_2 , . . . given beginning sets. The whole B_1 , B_2 , . . . sequence could be called a prediction.
The elements of each B_m are a kind of "or" because at least one must happen, but the B_1 , B_2 , . . . are meant in the "and" sense, because each must come through.
So this prediction concept has an abstract flair to it too. We might even think that it trivially implies the old obeying for the total $B_1 \cup B_2 \cup \ldots$
Indeed, if every B_m has a beginning from s, then together they will contain infinite many of s. But this was a mistake! We could have only finite many $b_1, b_2, \ldots, b_n$ beginnings of s, if some of them repeat infinitely in different B_m-s. The only sure way for infinite many beginnings of s appearing in the B_m-s is if they contain longer and longer ones from s. A simple condition on the B_m-s that assures this for any s, is if the B_m-s already contain longer and longer beginnings.
This means $\min B_m \to \infty$. Here $\min B_m$ denotes the shortest length in B_m and $\to \infty$ means that the values stay above arbitrary big N number, after an M index, that is $m > M \Rightarrow \min B_m > N$. We'll call this condition as the prediction being monotone. A more "subtle" condition that nevertheless implies monotony is diminishing: $< B_m > \to 0$. Indeed, having one n long beginning in B_m already gives $\frac{1}{2^n}$ value, so then $< B_m > \geq \frac{1}{2^n}$. Thus, if $< B_m > \to 0$, then only longer and longer elements can be in the B_m-s. So, monotony or diminishing of a prediction implies that obeying the prediction also means obeying the total as a set.
This result at once gives a fairly obvious equivalent of the $< B >=$ finite strangenesses among predictions. Namely:
s obeys a B that $< B > =$ finite $\Leftrightarrow$ s obeys a B_1 , B_2 , . . . that $< B_m > \to 0$
From B, with $< B > =$ finite, we get the B_1 , B_2 , . . . as follows: $B_1 = B$.
Then omit the 1 long beginnings, that is 0 or 1 if it was in B_1 to get B_2.
Then omit the 2 long ones, like 0 0 , 0 1 , 1 0 , 1 1 to get B_3. And so on.
In other words: $B_m = B^m \cup B^{m+1} \cup \ldots$
This is diminishing because the left out elements total chance value, that is
$< B^1 \cup B^2 \cup \ldots \cup B^{m-1} >$ approaches $< B >$, so the remainings' total, that is $< B_m >$ approaches 0.
Since $B_1 \supseteq B_2 \supseteq \ldots$ is finitely narrowing, thus an s obeying $B = B_1$, will obey all the members, which implies obeying the prediction too.
In reverse, from a B_1 , B_2 , . . . with $< B_m > \to 0$, we simply leave out the members until one has a total under $\frac{1}{2}$. Then leave them out till one is under $\frac{1}{4}$, and so on. The remaining prediction is clearly fast diminishing in fact its total set has a total chance value under 1. An s obeying the prediction will obey the left subset too, and since this is monotone too, s will obey their total which has finite chance total.
There is a natural, weaker requirement than narrowing, that is crucial for predictions.
The trick is to generalize the concept of "continuation" from beginnings to beginning sets. This is especially due because we didn't even introduce continuation of

beginnings officially yet. We used $b > s$ for b being a beginning of s, and I didn't explain why I used the bigger symbol for this. In fact, one might say that a beginning is smaller than a whole sequence, so the opposite smaller direction should be more logical. Well the logic is that we always should think about sequences. So a b beginning is "actually" the collection of all those sequences that can continue b.
We can even use [b] for this actual set and then s being a continuation of b means that $s \in [b]$ or $[b] \supset \{s\}$. Then, b being a beginning of c, that is c being the continuation of b is again logical as $b > c$ meaning $[b] \supset [c]$. In fact, here we can have $b \geq c$ too, for $[b] \supseteq [c]$ that is allowing $b = c$.
From all this now, the generalization to beginning sets is easy.
$B_1 \geq B_2$ simply means that every element of B_2 is either in B_1 or is a continuation of a B_1 element. So $B_1 \supseteq B_2$ trivially implies $B_1 \geq B_2$.
$B_1 > B_2$ means that every element of B_2 is a continuation of an element in B_1.
We might think then, that the same sequence logic follows to beginning sets too, that is if [B] denotes the sequences that are continuations of some elements in B, then $B_1 \geq B_2$ and $B_1 > B_2$ are merely saying $[B_1] \supseteq [B_2]$ and $[B_1] \supset [B_2]$.
The second strict continuation is clearly false this way because we can drop elements from B_1 and obtain the strict containment without continuing elements at all.
This suggests a very smart alternative concept to continuation, namely "covering". For beginnings they are the same, but for beginning sets we use the word cover for the [ ] containments, while the continuation for the way we defined $>$ above. Then we can see that cover always follows from continuation but strictness and reversal doesn't. To see that even non strictness doesn't reverse, observe that {00 , 01} covers {0} but clearly {0} is not the continuation of {00 , 01} instead this is the continuation of {0} and as it should, {0} covers it too.
A $C_1, C_2, \ldots$ is cover of $B_1, B_2, \ldots$ if each C_m covers each B_m, that is $[C_m] \supseteq [B_m]$. This at once implies that if an s obeys $B_1, B_2, \ldots$ then it will obey $C_1, C_2, \ldots$ too. In fact it will obey any $D_1, D_2, \ldots$ subset of this cover too. So we can produce single B strangeness from them too.
Of course monotony is still needed to guarantee obeying the total.
Thus we want a new requirement that replaces monotony and implies the diminishing of covers too. The idea is very simple! Just as the "external" $<B>$ = finite condition is better expressed by saying that the group values of B diminish fast, we look for the group values of the members in a prediction too. Then we can change monotony to a more probabilistic feature about these B^n_m groups.
But first quite generally, we say that:
B has q as portion bound if all $< B^n > \leq q$.
This $< B^n >$ is a probability, so could be denoted as $| B^n |$ too, because the total of the $| b | = \dfrac{1}{2^{<b>}}$ values for a fix group, if we regard all possible beginnings in it, is 1.
Indeed, if $< b > = n$, then we have 2^n possible beginnings and each has $\dfrac{1}{2^n}$ chance.
But for a whole B, the $| B^n |$ is not really a chance of having a beginning in B^n.
Not if we pick randomly from B, only if we pick from its n long beginnings.
So the portion bound is a more appropriate name and it simply emphasizes that all group's portion in B is under q. Of course, $q = 1$ is a trivial bound for any B.
The lowest bound is denoted as q (B), the quota of B. We could call this the maximal group portion of B too, but since we have infinite many groups, it doesn't have to be an actual portion value in any of them. It may only be the limit of them.

Now we view the $B_1, B_2, \ldots$ predictions so that every B_m member in it, is partitioned into its $B_m^1 \cup B_m^2 \cup \ldots$ groups.
Thus the quotas of the members can be regarded too. As m increases, we want to be more and more accurate in all the groups. In short, we want $q(B_m) \to 0$.
First of all, this implies monotony at once. Indeed, having even just one member in B^1 is $\frac{1}{2}$ portion, in B^2 it is $\frac{1}{4}$, in B^n it is $\frac{1}{2^n}$.

So if the quota is $\frac{1}{2^n}$, then upto n, we can't have beginnings at all.

Thus, the condition of a diminishing quota is much stronger than monotony.
It limits the number of beginnings with all possible n lengths. Namely, with a common portion bound. This maximum q portion is from 2^n possible beginnings in the n long group. So, $q\,2^n$ directly gives a bound on the number of beginnings with n length. $< B^n > \le q$ or $\text{num}\,B^n \le q\,2^n$ means the same. Indeed:
$< B^n > = \frac{\text{num}\,B^n}{2^n}$ because every member of B^n contributes $\frac{1}{2^n}$ to $< B^n >$.
Even though $q(B_m) \to 0$ is much stronger than monotony, it is also trivially weaker than the $< B_m > \to 0$ diminishing. Indeed, this implies it at once, since:
$< B > = < B^1 > + < B^2 > + \ldots > q(B)$

Lets pick a fix c. All possible continuations of it as a set is denoted as c^∞.

This set is covered by $\{c\}$. If c is n long, that is $< c > = n$ then c^∞ won't contain any $1, 2, \ldots, n-1$ long beginnings, so $(c^\infty)^k$ = empty for $k = 1, 2, \ldots, n-1$.

$(c^\infty)^n = c$, $(c^\infty)^{n+1} = \{c\,0, c\,1\}$, $(c^\infty)^{n+2} = \{c\,0\,0, c\,0\,1, c\,1\,0, c\,1\,1\}$ and so on. Observe that $|\,c\,0\,| + |\,c\,1\,| = \frac{1}{2^{n+1}} + \frac{1}{2^{n+1}} = \frac{2}{2^{n+1}} = \frac{1}{2^n} = |\,c\,|$.

This remains the same for all $(c^\infty)^k$ group for $k > n$ too.

$< c^\infty > = < (c^\infty)^n > + < (c^\infty)^{n+1} > + \ldots = |\,c\,| + |\,c\,| + \ldots = \infty$.

For a whole C set, we can similarly regard C plus all continuations from it as C^∞.
Clearly, this is "even more" ∞. From these infinites, it's obvious that a $< C_m > \to 0$ cover can't usually inherit its diminishing to the covered continuations, where all $< B_m >$ can be even infinite, no matter how small the $< C_m >$ were.
Thus, it's truly surprising that: If $C_1, C_2, \ldots$ covers $B_1, B_2, \ldots$ then

$$< C_m > \to 0 \text{ (fast)} \quad \Rightarrow \quad q(B_m) \to 0 \text{ (fast)}$$

Most surprisingly though, this has nothing to do with predictions because simply:

$$C \text{ is a cover of } B \quad \Rightarrow \quad < C > \ \ge \ q(B)$$

It's enough to show that every group's portion in B can't be more than $< C >$.
Those b elements in the group that are covered by same or earlier group elements of C, are okay because these early elements of C have a total chance value at least as big as the total of these b-s. The remaining b elements of B, that are covered only by later elements of C, can be replaced first by $\{b0, b1\}$, then by $\{b00, b01, b10, b11\}$ and so on. These all total the same in chance as b and all these elements must be covered by C elements gradually. C might even cover some branches more times, but just leaving out the first ones, we clearly have at least as big total in C as these elements amount to.

A perfect exercise to see quotas is our next example, showing that the reverse of this result can't be true. Indeed, we create a $B_1, B_2, \ldots$ prediction so that $q(B_m) = \frac{1}{2^m}$ but $< B_m > = 1$ and no cover of it can be smaller either. I will list the B_m-s quite explicitly but put the same grouped members under each other:

$$B_1 = \{0, \begin{array}{l} 10 \\ 11\} \end{array} \qquad B_2 = \{\begin{array}{llll} \overline{00}, & \overline{01}0, & \overline{10}00, & \overline{11}000 \\ & 011, & 1001, & 11001 \\ & & 1010, & 11010 \\ & & 1011, & 11011 \\ & & & 11100 \\ & & & 11101 \\ & & & 11110 \\ & & & 11111\} \end{array}$$

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

As we see, the first group in B_m that has member is B_m^m and the last is $B_m^{m+2^m-1}$.

$$\text{So, } B_m = \underbrace{B_m^1 \cup \ldots \cup}_{\text{empty}} B_m^m \cup \ldots \cup B_m^{m+2^m-1} \underbrace{\cup \ldots}_{\text{empty}}$$

And, $< B_m > = \frac{1}{2^m} + \ldots + \frac{1}{2^m} = 1.$

To see that no smaller C can cover any of these B_m-s observe that every sequence can be continued from B_m already and 1 is the minimal total to achieve this.

This prediction of course, was not narrowing. So we might think that narrowing would make such impossible. Not so, we just have to be a bit trickier.

Instead of starting for B_m at the m-th group, now we start much later, only after the next combinations are finished. This way we can include all of them together in B_1, then from the third group in B_2 from the seventh in B_3 and so on. Thus, B_m starts in the $2^m - 1$ group. Again, we'll list the same grouped combinations under:

B_1		B_2 starts here				B_3 starts here
0 ,	1 0 ,	$\overline{0\,0}\,0$,	$\overline{0\,1}\,0\,0$,	$\overline{1\,0}\,0\,0\,0$,	$\overline{1\,1}\,0\,0\,0\,0$,	$\overline{0\,0\,0}\,0\,0\,0\,0$
	1 1	0 0 1 ,	0 1 0 1 ,	1 0 0 0 1 ,	1 1 0 0 0 1 ,	0 0 0 0 0 0 1
			0 1 1 0 ,	1 0 0 1 0 ,	1 1 0 0 1 0 ,	0 0 0 0 0 1 0
			0 1 1 1 ,	1 0 0 1 1	.	.
				1 0 1 0 0	.	.
				1 0 1 0 1	.	.
				1 0 1 1 0	16	32
				1 0 1 1 1	elements	elements

$\overline{0\,0\,1}\,0\,0\,0\,0\,0$,	$\overline{0\,1\,0}\,0\,0\,0\,0\,0\,0$,	$\overline{0\,1\,1}\,0\,0\,0\,0\,0\,0\,0$,	$\overline{1\,0\,0}\,0\,0\,0\,0\,0\,0\,0\,0$
0 0 1 0 0 0 0 1 ,	0 1 0 0 0 0 0 0 1 ,	0 1 1 0 0 0 0 0 0 1 ,	1 0 0 0 0 0 0 0 0 0 1
0 0 1 0 0 0 1 0 ,	0 1 0 0 0 0 0 1 0 ,	0 1 1 0 0 0 0 0 1 0 ,	1 0 0 0 0 0 0 0 0 1 0
.	.	.	.
.	.	.	.
.	.	.	.
64 elements	128 elements	256 elements	512 elements

			B_4 starts here
$\overline{1\,0\,1}\,0\,0\,0\,0\,0\,0$,	$\overline{1\,1\,0}\,0\,0\,0\,0\,0\,0\,0\,0$,	$\overline{1\,1\,1}\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0$	$\overline{0\,0\,0\,0}$....
1 0 1 0 0 0 0 0 1 ,	1 1 0 0 0 0 0 0 0 0 1 ,	1 1 1 0 0 0 0 0 0 0 0 0 1	
1 0 1 0 0 0 0 1 0 ,	1 1 0 0 0 0 0 0 0 1 0 ,	1 1 1 0 0 0 0 0 0 0 0 1 0	
.	.	.	
.	.	.	
.	.	.	
1028 elements	2056 elements	4112 elements	

Here obviously $B_1 \supset B_2 \supset B_3 \supset \ldots$

In B_1, the first and second group has $\frac{1}{2}$ portions, the third upto the sixth, has $\frac{1}{4}$, the seventh upto the fourteenth, has $\frac{1}{8}$, and so on. Thus, $q(B_1) = \frac{1}{2}$.

In B_2, the portions start with $\frac{1}{4}$, then $\frac{1}{8}$ and so on. Thus, $q(B_2) = \frac{1}{4}$.

Then, $q(B_3) = \frac{1}{8}$ and so on.

In spite of the diminishing quotas, every possible s sequence has one of its 1 or 2 long beginning in B_1, then one of its 3 to 7 long one from B_2 and so on.

But also observe that with our method, the diminishing of the quotas became slow.

On the other hand, we achieved much more than just a narrowing prediction, that is obeyed by every s sequence, because of the strict form of obeying.

So not so surprisingly, even a narrowing prediction with fast diminishing quota, can be obeyed by all s sequences. The reverse of our earlier theorem seems impossible.

We created four "levels" of smallness about the quotas of the $B_1, B_2, \ldots$ predictions. Simple $q(B_m) \to 0$, $q(B_m) \to 0$ plus $B_1 \supseteq B_2 \supseteq \ldots$, $q(B_m) \to 0$ fast or finally $q(B_m) \to 0$ fast plus $B_1 \supseteq B_2 \supseteq \ldots$

We didn't examine all of these and how they relate exactly, but obviously the last is the strongest and we made it quite convincing that it is still not enough to be a strangeness always, that is to exclude random sequences.

What else can we do? Our last example actually hid the solution.

Indeed, the reason, that that example pushed forward the groups appearing in B_m way ahead of m, was that we included a lot of continuations of already appearing beginnings in a single B_m. Just look at 0 in B_1, $0\,0\,0$ in B_2, $0\,0\,0\,0\,0\,0\,0$ in B_3.

So having continuations in one B is necessary, so that sequences can obey B at all.

In a prediction, the B_m-s themselves don't have to contain continuations and our first example was such. Of course, narrowing predictions must contain continuations.

Requiring even more continuations than to achieve narrowing is thus, an even better method of leaving less room for new beginnings, and thus, to restrict the obeying sequences. So then, going all the way and requiring that the B_m-s contain all their continuations, is the way to go. In short, we'll require that $q(B_m^\infty) \to 0$ (fast).

Remember that the infinity power meant all continuations added to the set. We could avoid the use of this by simply saying that a set is "continuing" if all continuations are included. Then we would simply require that the B_m-s are continuing. This naming would then mean that a prediction is continuing if all elements are. This is illogical though, because a $B_1, B_2, \ldots$ continuing should mean what it meant already that the members are continuing each other. So instead of "continuing", a set should be called full or complete when it contains all continuations. Complete is the better choice, because it is a verb too, so B^∞ can be referred to as the completion of B.

A B is complete if $B^\infty = B$. A prediction is complete if all B_m-s are complete.

By the way, the official name for this complete feature among predictions is being "sequential", which doesn't reflect its meaning at all.

The use of ∞ power for completion is very visual too, as the one opposed to the $B^1, B^2, \ldots, B^n, \ldots$ fix length groups of a B. The ∞ is going beyond, namely that it is not a subset, rather adding new elements to B. A cute example of using the ∞ power is that $\{0, 1\}^\infty$ denotes the set of all possible beginnings.

The way we used the $1, 2, 3, \ldots \infty$ powers, suggests that using 0, that is B^0 should also be something opposed to both the groups, but most importantly, opposed to B^∞ too. Well, B^∞ is complete, that is contains all continuations, so B^0 should be opposite, that is not containing any two b, c that continue each other.

But how should it relate to B itself? Since B might contain already continuing elements, how should we choose the ones to be in B^0?

The solution is easy, we should choose the shortest ones. So to be precise:

A b is minimal element in B if b is not a continuation of anything else in B.

$B^0 := \{b \; ; \; b \in B \wedge b \text{ is minimal in } B\}$ = minimal subset of B.

There are three important features of B^0:

1. B^0 has no continuing elements.
2. $\langle B^0 \rangle \leq 1$.
3. B^0 is a cover of B and it is minimal, that is for any C cover, $\langle B^0 \rangle \leq \langle C \rangle$.

1. is trivial and we might think that this property itself could be best called as being minimal. But in a set without continuing elements all elements are maximal too, so we could call it maximal too. So the best name is non-continuing. Earlier I fought against the use of "continuing" for what I call complete and now I accept the opposite. In fact my argument was that for predictions, the continuing is already clear as continuing member by member. So then how should we call a prediction with all non-continuing members? Well simply as having non-continuing members. So then a continuing prediction with non-continuing members is quite clear and possible. In fact we'll see, that these play crucial role if the continuing is strict and then a new name will be "perfect" anyway.

2. is far from trivial. Indeed, B^0 can contain infinite many elements.
   Yet the proof is quite easy if we split up B^0 into its groups, that is as
   $B^0 = B^{01} \cup B^{02} \cup \ldots$ and check the maximal possible number of elements:
   num $B^{01} \leq 2$ obviously.
   num $B^{02} \leq 4 - 2$ num B^{01}.
   Indeed, in theory, B^{02} has four possible elements, but every element appearing in B^{01} blocks out 2 in B^{02}. Then again:
   num $B^{03} \leq 8 - 4$ num $B^{01} - 2$ num B^{02}.
   Indeed now, every element in B^{01} blocks out 4 and in B^{02} blocks out 2.
   In general:
   num $B^{0n} \leq 2^n - 2^{n-1}$ num $B^{01} - 2^{n-2}$ num $B^{02} - \ldots - 2$ num B^{0n-1}.
   So, $\dfrac{\text{num } B^{0n}}{2^n} \leq 1 - \dfrac{\text{num } B^{01}}{2} - \dfrac{\text{num } B^{02}}{4} - \ldots - \dfrac{\text{num } B^{0n-1}}{2^{n-1}}$
   Or taking all fractions from the right to the left:
   $$\sum_1^n \frac{\text{num } B^{0k}}{2^k} = \sum_1^n < B^{0k} > \leq 1.$$
   Of course, thus, $\sum_1^\infty < B^{0k} > = < B^0 > \leq 1$ too.
   Before, the total chance values for the groups could not be interpreted as chances. This total can be. Indeed, now we can pick randomly an s sequence by picking all digits randomly, that is perform a coin flip sequence and then $< B^0 >$ is the chance that the sequence has "at least" one beginning from B^0. The "at least" of course has no real meaning, because the non-continuing of B^0 guarantees that maximum one beginning can be from B^0. This shows that a certain artificialness is still hiding behind this seemingly perfect chance value too.

3. Clearly, B^0 is a cover of B and all other C covers cover B^0, so it's enough to show that a cover of a non-continuing set is at least as big in chance total. Indeed, those non continuing elements that are covered by beginnings are maximum as big in total and the ones covered by continuations again.

A trivial application of the minimal subset is that we can claim our earlier $< C > \geq q(B)$ inequality for the minimal subset as cover: $< B^0 > \geq q(B)$

The consequence for predictions is: $< B^0_m > \rightarrow 0$ (fast) $\Rightarrow$ $q(B_m) \rightarrow 0$ (fast)
We saw that the reverse didn't stand, by merely using narrowing predictions, and figured that complete members could be the way to go. Now it's easy to see why:

Simply because again it doesn't depend on predictions. For complete B in general we have a perfect equality: B is complete $\Rightarrow$ $< B^0 > \; = \; q\,(B)$.

Indeed, every b element of the minimal set can be replaced again by the continuing sets and these will have the same chances as b. The later and later groups will contain more and more of these disjoint continuations and thus will approach the total of the minimal set.

Finally, we can put our new result in a form using ∞ powers, and thus avoid the condition of completeness too: $< B^0 > \; = \; q\,(B^{\infty})$

So the two extreme powers provide an equality for any B set.

This of course means equivalences of diminishings for B_1, B_2 , . . . predictions:

$$< B^0_m > \; \rightarrow \; 0 \text{ (fast)} \quad \Leftrightarrow \quad q\,(B_m) \; \rightarrow \; 0 \text{ (fast)}$$

Most beautifully, an s obeys B_1 , B_2 , . . . if and only if, it obeys the B^0_1, B^0_2 , . . . or also the B^{∞}_1, B^{∞}_2 , . . . predictions.

But some things are not so beautiful: First of all, we have the fast options.

Are they necessary or merely extra to be a strangeness?

In fact, how do we find the old equivalent singular $< B > =$ finite strangeness?

This raises the question, why is continuing immaterial in our beautiful equivalence?

4. Ghosts of the Machines

We don't need the exact definition of machines. The reason for this is something quite amazing that somehow missed the attention of all the little foot soldiers in the new army of darkness. I refer to the computability math that overgrew even new math like a weed. They all do the obligatory homage to Turing and define his machines, these seemingly alien entities. And then they start doing math in a new way, applying the effectivity of the machines to some sets and thus, restricting the old math arguments or giving new twists to them. What drives these new arguments is not the actual usage of the machines, rather a common plausibility of what is effective and what is not. Now there is nothing new about this in itself.
The axioms of geometry are also just listed as trophies and then nobody thinks about them while solving problems. Of course, we can always dissect an argument and show that indeed, only the axioms were used. Similarly, in every other field too, the arguments can be turned to exact proofs, upto the level of effectively showing how the axioms were applied. So every math is relating to machines in this sense of derivability from the axioms. There is a little ambiguity here though:
Can machines derive the theorems or merely they can verify if a derivation is correct?
This ambiguity stems from the very natural experience that new proofs are based on new introduced concepts, new visions, new definitions. That's what gives math its organic depth. We can venture in it to explore new tracks. And yet, we solve old problems. Sometimes we even try to eliminate the new tracks and rephrase them into the old problem. This hiding is not healthy, it's not even good derivation. Most importantly though, even the new definitions are just formalizable properties or sets by exact rules. So in theory, we could try out all possible venturings into new tracks. The length of the exact forms are restricting the possible variations, so going through longer and longer hypothetical definitions, we can go through them as a sequence. We can't finish them all, but we don't have to. We can merely try them, one by one, where they lead to. Of course, where they lead is again an infinite possibility. So what we have to do is what we indeed do in practice, namely trying out the alternative paths, simultaneously. More and more options, checked for more and more possible consequence. It's still just an infinite sequence of roads. If one of them yields a solution to an old problem, then we succeeded.
So in a sense, discovering derivations for theorems by intuitions or mechanically, is not incompatible. And indeed, as computers became faster, the chess machines became better too. They search blindly for the best moves by merely checking all the possible pathways. The champions venture ahead by intuitions. Everything we discover in math can be shown later to be potentially discoverable by a machine too. The only exception is if we have to introduce new axioms. A machine can only use the ones we programmed in.
The fact, that we can define machines and give its rules as axioms, already raises an interesting paradox. Will the consequences be strange, if the axiom system talks about itself? Most importantly though, the mechanicalness of effectivity can be achieved in many ways, so even axioms that don't talk directly about derivations, might exactly spell out the rules of all derivabilities. So the system can be conscious of derivations, but not self conscious about this. Then, the same strange consequences must apply to these "naïve" systems. This is the exciting second side of Logic.
But the new math of computability is not this. Here the derivability itself, is not observed. Instead, objects that are effective are examined. So this could have been done already before the whole self examination of Logic started. Indeed, it's a mystery why it didn't happen. Most importantly though, we don't even need the

axioms of effectivity, to do that. Just as one doesn't have to know the axioms of geometry, to do geometry. The late emergence of this effective math explains its inferiority complex and thus, its blatant Formalism. It simply hasn't even gone through a healthy naïve period.
I try to do my best in the followings to help this sick "man-child" and let him play.
All math only comes alive if we play with it. The intuitions of machines are in all of us. We don't need the axioms yet, rather let the ghosts of the machines enter math.
Of course, our fix points for using machines were the beginning sets. We knew that infinite sequences are beyond machines. That's why we turned to beginnings. We also hoped that using machines to define beginning sets will overcome the problem of allowing all sets of beginnings. Remember, that this would be a failure towards defining strangeness. Indeed, for any s random sequence, the beginnings of s is a very small set. It would be very strange to have infinite many from these. In fact, it implies the exact repetition of s. So it's a strangeness. And yet, s obeys it, so it's not a strangeness. The error was that we employed a hypothetical random s. But how can we avoid this? Our hope is the machines. This is the qualitative finiteness. A machine can't produce exactly the beginnings of a random s, that's obvious. But we have to lock out every such concrete random sequence. And indeed, obeying means only having infinite many from B. How can we exclude this? The solution was the quantitative finiteness of $< B >$
The reason we were not hundred percent happy with this, was that the $< B > =$ finite rule was more an avoidance of $< B > = \infty$, which we showed pretty convincingly, to allow random s sequence "most of the time". So $< B > = \infty$ could still contain some strangeness too, that is exclude randomness sometimes.
So we defined a new form of strangeness as picking not from a single pool of B beginnings, rather from a sequence of $B_1 , B_2 , \ldots$ one by one. We called this a prediction. We showed that single B can produce $B_1 , B_2 , \ldots$ and vice versa.
We can regard any B set, partitioned into its different long beginnings that is groups: $B = B^1 \cup B^2 \cup \ldots$ With this $< B > =$ finite becomes $< B^n > \to 0$ fast.
In fact, it's actually a definition of fast diminishing, so we merely gave a new vision of the same rule. $< B^n > \to 0$ fast is more "internal" condition than $< B > =$ finite.
But for predictions the diminishing of a cover doesn't have to be fast because any subset can be chosen as new prediction, so clearly we can get a fast diminishing anyway. Apart from this weirdness, we tried the partitioning for predictions too, to replace monotony. This was a full success only after the heuristic definitions of:
$q(B)$ the quota of a B is the upper limit of the $< B^n >$ group "chances".
B^{∞} is the completion, that is adding all continuations to B.
B^0 is the minimal subset, that is only keeping the shortest from continuing ones, that is keeping only the minimal elements. This is the minimal cover of B too.
Indeed, we have for any B: $< B^0 > \; = \; q\,(B^{\infty})$
This implies the equivalence of the diminishing of these two for predictions too.
So now, the diminishing of the covers can be established internally too by the diminishing of the quotas if the prediction is complete $B_m^{\infty} = B_m$.
Continuing doesn't matter here but it was necessary to obtain single B strangenesses. Could this be the clue to refine strangeness among predictions?

Trial sequences had a false timely meaning already and now machines bring in a new time illusion of their own. So first lets get rid of these.
The trial sequences can be imagined in space. Even the word prediction used for our alternative strangenesses should be visualized in space rather than in time. It merely means boxes from which the sequence of beginnings must pick one from each box.
The picks don't have to be continuing by definition but if the prediction is monotone that is the boxes contain longer and longer beginnings, then eventually the picks have to be longer and longer too. Then having a pick from each box also means obeying the total of the boxes, that is having infinite many beginning from it.
Continuing prediction means that all the beginnings contained in any box are either already in the previous or continue one in the previous. So, beginnings that are not continuations of any other must be listed already in the first box. The second box can contain some of these again but new ones must be continuations only. The third again some old ones or new continuations of the second ones, and so on.
This continuingness with monotony implies the crucial property of obeying the total.
There are four specializations of the continuing predictions.
Strictness means that we don't allow repetitions. All elements are continuations of previous ones. An other way of saying this is that the boxes are disjoint.
Non continuing members, that is boxes, means that there are no continuing beginnings in them. These two special features can be combined as "perfect":
Such strictly continuing prediction with non continuing members is a perfect prediction in the sense that it only allows continuing beginnings to be picked. This is what we would imagine as an actual prediction in time. Also, in such perfect prediction we actually sorted the predictable beginnings according to how many predictable beginnings they have. Indeed, the first box contains all the ones that are minimal, that is have no predictable beginning. The second contains the ones with one predictable beginnings, and so on. So these boxes are determined by the total B.
In fact we already denoted the first box as B^0, the minimal subset of B.
We can't continue this notation because we used the natural exponents for the length partitioning of B, that is the length groups.
Using bracketed exponents, we get this new, "perfect" partitioning of B as :
$B = B^0 \cup B^{(1)} \cup B^{(2)} \cup B^{(3)} \cup \ldots$
Totally opposite directions of the continuing predictions are the following two:
Narrowing means that the boxes contain all later ones, so we list all beginnings already in the first, then leave out some again and again. The advantage of such predictions is that a sequence picking a beginning from a box, at once picks it from the earlier ones too. So picking from arbitrary late box means obeying the whole prediction. Thus the obeying of the total can be forced to mean obeying the prediction with some added conditions. An other way of looking at this, is that obeying the total means obeying the first box and we can force this to the next, and so on. We used this with finite narrowing. Then clearly a sequence that obeys the first box must still obey the next, and so on. The above defined perfect partitioning of a B can be used to define an also "perfect" narrowing prediction, by starting with the B total and dropping out the perfect groups. First the minimal beginnings, then the ones with one sub beginning, and so on. This is not finitely narrowing and yet we get the same effect. A sequence obeying the first member must obey the next, and so on.
Finally, the fourth special continuing predictions are the complete ones, meaning that in any box we include all continuations too. This makes sense too, because if a continuation becomes correct it means the success for the shorter one anyway. These were used for the concept of quotas but they can be important for machines too.

Using a machine to produce a single B beginning set or a prediction sequence seems to suggest that we at once restricted the possible continuations. But this is false.
We can generate all beginnings as: 0 , 1 , 00 , 01 , 10 , 11 , . . .
Even create a perfect prediction by boxing them as the increasing length ones.
That's why we needed some quantitative restrictions beside the qualitative finiteness by machines. This was having a finite total for single B, or having diminishing cover or quotas for predictions. These exclude bundles of random sequences going through as realities. But the beginnings of individual random sequences define narrow beginning sets that they themselves will obviously obey. These sets of course are purely theoretical because we cant give a concrete random sequence.
To overcome this problem, we have to shift from the non concreteness of the random sequences to a concreteness of the beginning sets.
Machine generation of the beginning sets is exactly doing this.
Machines are the qualitative finiteness to exclude all random individual sequences, that fool the quantitative restrictions.
Individual random sequences can't obey such concrete narrow beginning sets.
They can obey concrete bundles or narrow abstractions but not concrete narrows.
Clearly we want to use this idea the least restrictively as possible. And indeed, there is a widest meaning of beginnings recognized and thus beginning sets collected by machines. This will also mean a possible listing of such beginning sets but not increasingly by their lengths. So the idea of a machine predicting longer and longer beginnings is false. A short beginning might be listed much later than a long one. This seems stupid in a "casino environment" but this is the deeper picture. The outcomes are not in time, or if you want them in time, then wait till they are all finished and establish the strangeness for the whole sequence. This doesn't mean that later a strictly forward going, gambling definition can't be also given.
This use of machines for beginnings as objects to be recognized or produced, came from our goal to grasp strangenesses about sequences, but it turned out to be a perfect coincidence with machines in general. Indeed, machines can only deal with objects that have an internal finite structure and not with objects that are merely abstract elements of external infinite structures like numbers are. Numbers for machines, are merely the actual digits, just as words are the sequence of letters. We see this fusion of numbers and texts in actual computers. The memory of a computer is the crucial background needed for deciding facts about numbers or words. The big contradiction is that the memory if addressed by numbers, requires the very thing we wanted to avoid. Or if we address them by the actual numbers as words, then we need increasing lengths wasted. But there is a way out! Namely, we don't use addresses, rather access them step by step, going from fix positions only. Real computers do this by circulating long sequences of bits and grasping them by a clocking. An abstract computer could do it through purely spatial neighboring steps from bit to bit. This would be very slow in practice. But the point is that in theory, an infinite potential memory doesn't require addresses at all.
This infinite process of the alteration of the potentially infinite memory is merely the verification process about the initial memory or input. A condition about the process is what we require about the input. Of course this condition will only be true for some inputs. Importantly, even for these true or "yes" inputs, this condition might take arbitrary long to achieve. So doing the process only for a fix number of steps, we can not be sure yet of a yes or no answer. But most importantly this is still an asymmetrical situation because if yes happens then just going forward with the process, eventually we get an answer for all true inputs. So we can collect these true

inputs effectively. The negative inputs can never be collected because they still may come up as true.

So, the infinite verification process as Raw Effectivity is the basis of the consequential asymmetrical yes effectivity collection without a guaranteed collection for the no cases. For some yes collections of course we might be lucky and find a yes collection for the no-s too. These are decidable collections. But some effective collections don't have effective complement. This asymmetrical feature of effectivity is the same as the non increasing generability of the effective collections.

First of all a generation of a collection can go by going through all possible inputs that is beginnings. This seems contradictory because we need arbitrary many steps for each next input to be tried. But instead of trying one for ever, we can try it for one step. Then try a next input with one step. Then come back and continue the previous. Then try again a new and come back to both earlier ones. And so on, we can run more and more inputs longer and longer, so they all will be verified. But not in any fix order. Indeed if they could be verified in increasing order for example, then we could at once verify the complement by checking out the generated ones up to the same length.

So the use of machines means only recognizable, that is effectively collectable but not necessarily decidable beginnings.

Beside this fundamental input oriented use of the machines, they still produce outputs too as the alterations. In real computers, this over shadows the whole yes/no input meaning. The theoretical machines started to be used for the outputs, in order to grasp the concept of information. Indeed, a short input creating a long output means that this long object is compressible. For example an output containing a billion alternating digits can easily be produced by the rule itself, so it can be produced even from empty input. This shows that something is fishy here. Indeed, the machine itself can hide not only such methods but random memory too, so bigger and bigger machines can produce arbitrary big random objects from empty inputs. In fact it shows that inputs are not important at all. The collections could go for parts of the machines themselves. The use of inputs is merely a nice convenience to externalize the collectable parts.

But there is a secondary use of the input too. We can feed an other machine's crucial parts there and thus imitate that other machine. But importantly, using possible inputs for that other machine too, so using two inputs combined. This way the imitation doesn't have to be full imitation, which would be impossible for very large machines to be imitated. We only imitate the actual alteration process but take the data to be altered externally as second part of the input. That can be arbitrary long and so the imitator machine doesn't have to be arbitrary large. In fact single machines can imitate all others with this now crucial input usage. This then gives a solution to the information content goal. Namely the minimal inputs are the compressibilities if we stick to using a fix universal imitating machine. Indeed, this fix machine can hide only a finite many tricky complex outputs. Beyond a length, the minimal inputs to produce a wanted output will measure the compressibility of the output.

But when this ingenious new use of the machines succeeded, it needed additional restrictions about the machines themselves. So it is not a perfect coincidence!

Returning to our subject of defining beginning sets, we might expect that the crucial new problem brought in by machines can only be the use of predictions, that is a sequence instead of a single set. In a sense it's true but not in the way we would think. Indeed, a single machine can collect the whole sequence of sets.

But what is the recognition by? Giving the beginnings with a potential index as input or giving the beginnings and then getting a sequence of indices as result. Luckily, it doesn't make any difference. Indeed, suppose that a machine can recognize the beginning plus its index. Then an other machine can generate the beginnings also by simply trying out the beginning with the increasing numbers as potential indices, simulating the first machine and stop when one is successful. Of course again we have to use our earlier trick of step by step trials. In reverse if a machine can generate the indices then for an input of beginning plus index we can wait with the generation till it comes up. So if it's all that perfect then what new feature can come about?

To see this, lets try to follow through our simplest argument about the equivalence of the single B and predictions. From a finite total B, we omitted the longer and longer beginning groups to obtain a diminishing prediction. In reverse, in a diminishing prediction we omitted members to make it fast diminishing and then the total is B.

The first direction is machine friendly because the lengths of the beginnings are easy to recognize. The machine simply counts the length of the beginning we feed as input and checks the belonging to B. If this is yes then the length tells how far in the boxes it is a member. The reverse direction is problematic! To eliminate enough members, we have to go until a member's total is under some chosen value. It will happen with any values we wish. For example, we could select members with totals under :

$\frac{1}{2}$, $\frac{1}{4}$, . . . values and thus giving a full total under 1. But these members must be given effectively, in order to decide which beginnings to keep.

We can go through all boxes but we can't tell whether a given value is an upper bound. We could easily tell if a chosen value is not a bound because trying out elements and add their chances will then eventually overflow the chosen value. Rejecting such negative candidates still would leave us uncertain about which boxes to keep for sure.

So this at once suggests a new heuristic duality that replaces the old crucial distinction between fast and slow diminishing. Now it is "hard" versus "soft". Soft merely means a theoretical fact of diminishing, while hard means that there is an effective formula that gives the diminishing. This then easily allows a fast subsequence to be selected.

Now comes a second blow! Accepting this hard diminishing as the new condition for predictions, then the first easy part of our early argument will collapse. Namely, leaving out the longer and longer beginnings from a finite total B, does not provide an effective diminishing. First it seems to do so, by observing that now the passing of some bound in a group will guarantee that leaving out the whole group, the decrease is bigger than the bound and so the next box is less than the previous minus the bound. But this would also require chosen bounds that add up to the total of B. Otherwise we only get effectively decreasing boxes but not necessarily diminishing ones.

Luckily using our new "perfect" narrowing prediction that is narrowing by the,

B^0 , $B^{(1)}$, $B^{(2)}$, $B^{(3)}$, . . . groups, instead of the length groups, will work. We also change from diminishing to merely having diminishing cover, which we already explored. So the perfect condition for the predictions is having hard diminishing cover.

Before we show that this stands for the perfect narrowing prediction, I want to show an irrelevant but interesting twist. The, old length group narrowing didn't work due to the non guaranteeable hard diminishing, but the machine creation was obvious for the groups and so also for the narrowing boxes. A machine can select the lengths and create the remainings if increasing lengths are to be ignored. Here, the increasing sub beginning groups themselves are not machine recognizable. Indeed, we can't tell if a beginning might have more sub beginnings to be recognized after waiting longer.

Remember, that short beginnings can require long verification. And yet, luckily the narrowing boxes <u>are</u> recognizable. Indeed, these contain beginnings that have at least a given many sub beginnings, so once a new sub beginning is recognized, we can promote the beginnings into the next narrower collection too.

Now, turning to the hard diminishing cover:

First of all the perfect groups all cover each other: $B^0 > B^{(1)} > B^{(2)} > B^{(3)} > \ldots$

So, actually they all cover the totals from them too. But also, the total of the chance values up to the first n member is less than the total of B.
So the n-th member's total $<B^{(n-1)}>$, is less than the n-th of <B>.
So the narrowing boxes are all covered by these bounds too.
This saved the day and might suggest that it was meant to be that the old equivalence remains with the new effective meanings. Of course then we could ask what the whole purpose of predictions is. And I already foretold that they lead to the new alternative strangenesses. The crucial question is this: What happens if we keep the machine definition of predictions and also the diminishing cover but drop the hardness, that is the effectivity of the diminishing? Will it lead to obvious failure? That is, can we give sequences that are strange with this wider condition but shouldn't be strange intuitively? In a sense it is already clear that if we can chase such concrete cases then they are strange, so should be regarded as non random. But it could still happen that they are unspecifiable in any sense. Luckily they are describable quite well.