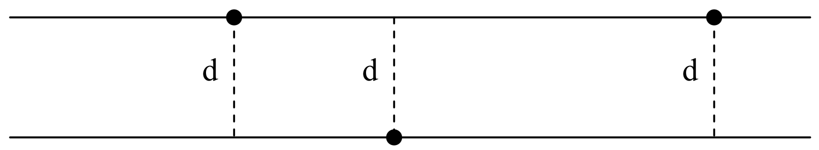
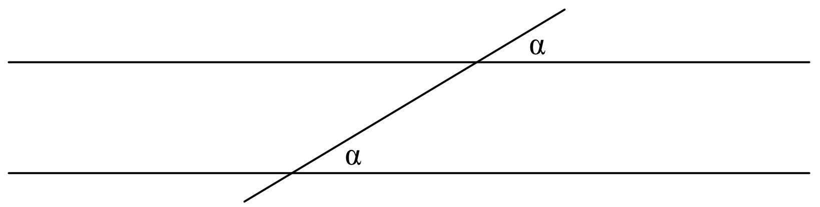
Plausibility Versus Logic

Proving things exactly, means that we use fix general steps as Logic and unquestionably true, plausible assumptions about our particular field as Axioms.
But not all plausible facts should be axioms because some plausibilities can follow from simpler and thus more plausible ones. In fact, these plausibilities have strange interrelations.
In Plane Geometry for example, the assumptions that two points determine a line and that two lines cross in a single point, feel as assumptions that couldn't be obtained from simpler ones.
And yet these two assumptions seem to imply each other indirectly. Indeed, if two lines had more than one common point, that is at least two, then two of these points determining a single line would contradict the two lines we started with. In reverse, if two points were not determining a single line, rather there were more so at least two going through them, then two such crossing in only one point would contradict the assumed two points we started with.
In spite of such nice "logical lines", it's easy to make mistakes. We made two above!
Namely, we jumped to the indirect negatives as "more than one" which means that we ignored the possibilities of "less than one" that is none. This would mean two lines not crossing at all and two points not determining lines at all. The first is indeed a case if the two lines are parallel but the second is an impossibility. This second still means we made a mistake because this impossibility must be claimed as an axiom. In other words we have to claim that any two points have a line through them. The determination, that is uniqueness of the line is an additional claim.
The first mistake of forgetting about the parallels is leading to a much deeper jungle.
Should we draw attention to such problems of plausibilities? I say yes!
We have to be "didactical"! Namely, paint a picture that shows how all common sense plausibilities relate to our claims. Even if it turns out that in the end our plausibilities are false or misleading. I claim that such didactical road always exists. The universe is didactical.
It is also "dialectical" or "dialethical". That is, it contains contradictions that come alive as reality and to our understanding as paradoxes too. But these themselves can be shown didactically.
Back to parallelity. It means three equally plausible appearances:

The first is "fix distance". Parallels are the same distanced everywhere. That is, all points of either one has the same fix d distance from the other line:



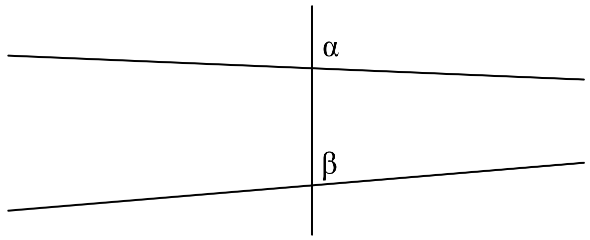
The second plausibility is "equi angle". Two parallels crossed with any third, will give same angles at the crossings on the same sides:



Finally, the simplest plausibility is "non crossing":

Euclid knew very well that these three are the everyday plausibilities of parallels but he never emphasized them because he was already entangled in the deeper questions of how they relate.
They imply each other. But this is not trivial at all.
If one of them would trivially imply the other two or follow from the other two, then probably this plausibility should be regarded as the definition of parallelity and the other two could become theorems about plausible properties or plausible conditions of parallelity. But neither is the case!
The only triviality is that fix distance implies non crossing.
So, these logical derivabilities override the mere plausibilities in the eye of the mathematician. This big trap is called formalism. The treatment of a field will become logical but not plausible.
It is not intentional confusion or hiding but a root of these. When the social acceptance of authority produces not elucidations rather further formalisms for its own sake as showing off then the "parrots" create Formalism. The "polishing" becomes intentional hiding, "overdoing" even the creators. The Wkipedia math articles are all such useless rubbish.

Euclid used the following axiom to prove the equivalence of the three plausibilities:
If we measure two non equal angles from a line on the same side and same direction, then these as lines will cross. Namely if we regard these angles as the right side and the upper one is the bigger, then on this right side but if the lower is the bigger then on the other left side:

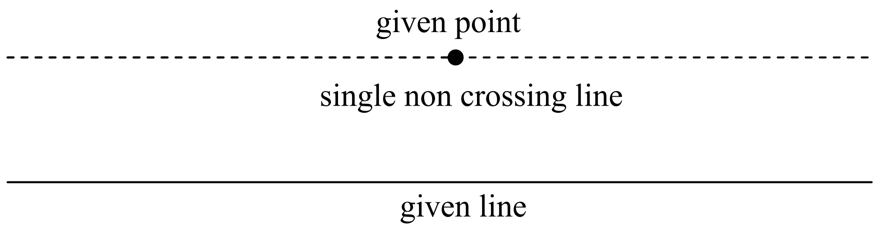


$\alpha > \beta$ so cross in this side

The specification about what side the crossing will happen is not crucial but even without that, this is an "ugly" axiom. The mathematicians who followed were not satisfied at all. Even two thousand years later, many still tried to avoid this axiom. Namely, tried to prove it as a theorem from the simpler assumptions of Euclid. Two results came out of these unsuccessful attempts.

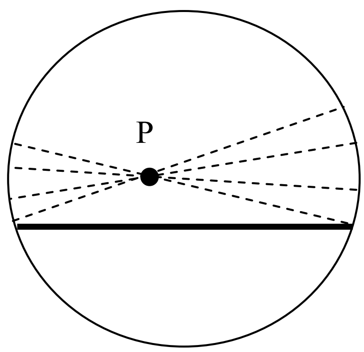
Firstly, a much nicer axiom that can replace Euclid's and this is what we teach in schools today:

For any given line through any given point outside the line:
There is only one line non crossing the given..



Indeed, this axiom talks only about non crossing and it doesn't involve distances or angles at all.
With this axiom of course it is almost obvious that a definition of parallelity should be the situation in the axiom, that is the non crossing. Then the real surprise is that this axiom will imply the seemingly unmentioned other two plausibilities, the fix distance and the same angle.
Eliminating this new axiom still remained a goal and these trials lead to deeper puzzles.
The method of elimination was to assume the axiom false and try to arrive at a contradiction.
Our so called "indirect" Logic would then accept this as derivation of the axiom itself.
But instead of contradiction very nice features were unfolding.
Janos Bolyai wrote it in his diary: From nothing I created a new different world.
His father was school mate of the famous Gauss and so Janos asked him to deliver his discoveries for evaluation. Few months later a short reply said that "I can not praise you, otherwise I would praise myself, since I arrived at the same result years ago".

Janos was shattered not without reason. The insensitiveness of Gauss was inexcusable but he wasn't lying. He put his earlier results into the drawer because he felt nobody would appreciate it.
The Russian mathematician, Lobachewsky saw the same new world and published this non Euclidian geometry first. Soon the whole world knew that unique parallels are not a logical necessity because their denial is far from meaningless. So what's going on?
Is logic more important than plausibility? Not at all.
This new world couldn't be interesting without new plausibilities.
So, then the unique parallels are merely a physical choice of plausibilities? Or maybe not even physical if we make more exact measurements? But what is a line physically anyway?
While all these philosophical problems were boiling, a lot of mathematicians realized that there are inside consequences of the non Euclidian geometries in math, avoiding physics or philosophy.
Namely, in our normal space or plane we can define new lines ourselves. Sometimes new distances and angles too, or keep one as the normal. But the resulting "model"-s will obey the axioms of geometry except for parallelity and thus "show" these new worlds inside the old.
The simplest one was doing a bit more than giving up single parallels, namely giving up parallels at all, in fact even the infinity of lines. This so called spherical geometry is the plane replaced by our earth's surface:
The shortest lengths between two points are the arcs of so called main circles having center in the center of the earth itself. Like the equator or the vertical time lines but they can be drawn slanted too, through any two points. The shortest flight from New York to L.A. is such. Going on such lines long enough of course we return to where we started, so indeed the lines are not infinite long.
Moreover, to keep the laws of two points determining single line and two lines crossing in single point, can only be achieved if we regard the diametrically opposite points as one.
So the north pole and the south pole, or opposite cities on the globe are merely "one points".
Even more surprisingly, there are no parallels! Any two main circles cross in two opposite points which as we just agreed are regarded as one.
The triangles have "curved" sides that are actually arcs from main circles.
The angle of two crossing arcs has a real meaning in the real space, namely as the angles of the tangential lines placed at the crossing point. These of course stick out from the surface. Yet these are to be regarded as our new angles as if they were inside the surface.
Thus the angles of a triangle will add up to more than 180^o.
A much more important, truly non Euclidian plane can be created in a circle if the perimeter is the infinity of the normal plane. So the distances must be distorted and as we go toward the perimeter, they are compressed more and more. Amazingly, all this can be done so that the angles remain the same and so the straight lines are strings of the circle. The multitude of non crossing bundle through a P point is directly visible:



So, these models were models for the axioms of Geometry except the parallelity axiom.

The Birth Of New Math And Gödel's Completeness Theorem

One such model explorer Beltrami, made a conclusion from all this, that with our hindsight could be called the exact beginning of "New Math". He realized that strangely, these non Euclidian models proved that Euclid was right! Not in his restricted view of parallelity but in his view that in this restricted world the parallelity axiom is unavoidable. So, those who tried to derive the parallelity axiom were in hopeless search and this can be proved now. But how?
Well, any logical derivation must be a true consequence in the models too. So since these strange models obey the axioms except for parallelity, if parallelity would logically follow from the others, it would have to be true in these models too. But as we see that it is not so.
So, Beltrami as the first in history, proved that something is not provable. And all this before even proofs were defined exactly. No wonder that from this point on, things accelerated. Sets were discovered by Cantor. These give the models. The rules of Logic were finalized by Frege and Hilbert. Beltrami's assumption that models obey logic became an almost triviality of sets.
The reverse question, whether Logic obeys Models is more tricky.
It would be obviously crazy to claim that if something is true in a model then it is logical too.
But if something is true in <u>all</u> models then it should be a mere logical necessity.
Applying this to a consequence claim $A \to C$ then, it should mean that:
If in all models where A is true C is also true, then Logic should be able to derive $A \to C$.
An additional next step of a good Logic should also be that instead of deriving this abstract consequence claim, it should also be able to derive through its rules the C consequence from starting with A. This we denote as $A \vdash C$. The consequence claim $A \to C$ can be generalized to more $A_1, A_2, \ldots, A_m$ assumptions if we introduce their "and" as new single claim. But the alternative, derivability from more assumptions is even better because this $\vdash$ is not a fix form rather a process and so we can claim even infinite many $A_1, A_2, \ldots$ assumptions or axioms.
The $A_1, A_2, \ldots \vdash C$ claim then means that there is a derivation of C from some finite many of the axioms. So, here too we need only finite many but we don't specify which. Even better, the perfection or "completeness" of a Logic can now be expressed by how the models must behave.
Namely, in every model where all the $A_1, A_2, \ldots$ axioms are true, the C consequence must also be true. To check this, would require to check all models obeying $A_1, A_2, \ldots$
Luckily, a much easier checking of this completeness of our Logic could be achieved.
Firstly, our logic is indirect. So, for C to be a consequence of some axioms, means that assuming the negative of C, that is adding $\neg C$ to our axioms, we must be able to derive a contradiction.
This contradiction implies that in all models of $A_1, A_2, \ldots$ the C is true, that is $\neg C$ is false.
This is still just the models obeying Logic and the reverse, that is Logic obeying the models means that if $A_1, A_2, \ldots, \neg C$ has no model then a contradiction is derivable.
Secondly, leaping from this $A_1, A_2, \ldots, \neg C$ to any sets of statements, we obtain that:
If any set of statements has no models, then our Logic should be able to derive a contradiction.
Thirdly as a final twist, regarding the negative version of this claim, we get the easiest criteria of a complete Logic as: Every set of statements that leads to no contradiction, must have a model.
With this, to prove the completeness of a Logic, we'll even have a practical method too.
Namely, we start from a hypothetical set that leads to no contradiction, and turn the unsuccessful contradiction searches into a model. So, the historical method of contradiction searches that lead to non Euclidian geometries came back as universal method. The difference is that while at the non Euclidian geometries it was the beauty that convinced us of a consistency first, and we found models later. Here in the general method, we create ugly models from assumed consistency. But this is only a new clear vision. In Gödel's original proof it was not just the models that were ugly but the road to obtain them too. So, Hilbert who finalized the rules of Logic, and was present at Gödel's presentation in 1930, was not really happy. Gödel didn't care much about this. In fact, he knew that the ugly models hide something much worse about axiom systems. He even hinted these ideas already at the conference informally, but probably only one Hungarian young man, Janos Neumann grasped the revolutional new ideas perfectly.

Our Grandest Silver Platter, The Sets

About the simpler non parallelity axioms of Euclid, I mentioned that already they follow a different logic in plausibilities than in derivabilities. And yet they seemed to survive unchanged till our present. Well, this is not true at all. Something happened in the twentieth century that we are not fully aware of yet. It has a prehistory and the year 1900 had a formal importance too.
At the Paris math congress Hilbert gave an important speech. Not only listed all the famous unsolved problems but gave a new direction of math. He explicitly emphasized Set Theory as the new paradise from which mathematicians can not be expelled ever. That was a nice gesture towards Cantor the inventor of Set Theory, who was still regarded as crackpot by some.
It’s a tragedy that he died in an asylum later.
Hilbert not only finalized the rules of Logic but went back to Euclid and reaxiomatized Geometry. Now, based on Set Theory. So Hilbert knew from this project concretely and personally what he preached, that Set Theory is the base of all math. This Set Theorization of Geometry on its first level meant that lines, circles are all point sets. The points themselves have no elements. They are atoms or urelements. When lines or circles cross then the crossing points are simply shared common elements. This vision became so obvious that today we can not even visualize these geometrical objects and their crossings in any other way. But all this happened definitely not due to Hilbert reaximatization. His system on its more important second level is so complicated and unnatural that it will never enter the education system. So what happened? We don’t know.
But the plot thickens! Set Theory disowned its lovechild, these urelements, and went in totally new directions. The stupid Venn Diagrams are the only tangible traces of the simple point set visions in elementary education. But they merely reinforce the obvious. That’s why they are stupid because they pretend to give something more. The real “more”, the equivalences and thus comparison of infinites are blocked out of elementary education. So, the story is not over yet.
But the other mystery is more than two thousand years old. Namely, why didn’t Euclid regard the lines and circles as point sets? This side of the puzzle contains the then still fully alive point paradoxes of Parmenides and Zeno. Today we “defused” some of these by two other silver platter concepts, the infinite decimals and the division algorithm.
So again, the first mystery is how the sets especially the point sets became a plausibility in the twentieth century and the second mystery is why they couldn’t become plausible before.
But the infinite sets have a third mystery themselves. But first lets recap things:
New Math started with Beltrami’s realization that reality obeys logic.
The reverse question that is logic obeying all realities means that our logic is complete.
But these realities are meant as set constructions that we called Models. If we start with such reality without an axiom system in mind to be “modeled” than we call these realities Structures.
They are the same things, sets representing both the objects and also the relations among them.
The crucial new axiom that came to light about sets, is the Axiom Of Choice.
The first formal importance of it is that it is actually a Logical axiom too. It simply generalizes the plausible rule of concretizations, that is using an already accepted name for a universality or introducing new name for an existence. Picking elements from accepted sets is a sample of sets using both universality and existences. This axiom is the most beautiful and solid point of both Logic and Set Theory. The perfect bridge between them. The idea that it is somehow a source of later paradoxes is insane. The real problem is that it is not enough, something else is missing.
A concrete sign of all this is how the plausibility is instantly shifted for both the Logical and Set Theoretical fundamental application. The fundamental application in Logic is the Completeness Theorem and indeed it is not the plausible Axiom Of Choice but its consequence the Wellordering Theorem that gives the plausible proof. In Set Theory the fundamental application is the comparison of sets and guess what. It is again the Wellordering Theorem that is the plausible cause. But unlike in Logic, here this comparison of sets leaves a big hole, so we are not really able to compare infinities. This hole is the undecidability of the Continuum Hypothesis.
I told all these in advance to see that Sets are not really a finished matter and so are a shaky ground for a Logic. In spite of the seemingly perfection of the Completeness of Logic.

Basics Of Logic

The essence of our Logic are the two quantors, existence and universality.
Amazingly, Aristotle formulated already this but in a bad way. What was the problem?
The seemingly simpler principle that was missed, is that all descriptions of a reality contain two different class of formal concepts, namely one relating to objects and one to states.
The states are the "state"-s that the objects can be. So while an object simply is what it is, the states are a yes or no possibility for them. Now, for the objects we have an easy duality again:
Talking about a fix one we don't actually grab it and show "this", rather use a "name" for it.
But when we talk about objects in general, then we use "variables" for them.
In everyday languages, the words for variables appear only as: somebody, everybody, something, everything. So, it is at once tied to the quantors of existence or universality.
This was the main cause of that variables were so hard to introduce and also that the two quantifications were fairly easy to realize but became falsely formalized by Aristotle.
The crystallization of variables came not from Logic, rather practical math, mostly algebra like equations. But two hundred years ago this itself was still in a messy confusion.
A state with one variable is called a property and is abbreviated as P (x). The variable is x, used in the P property. For more variables, we call these states as relations like R (x , y).
If we replace a variable by a name, or as I call it "concretize" it, then the state looses a variable.
For example, a two variable state becomes only a property. If we replace all variables by names, that is fully concretize it, then the state becomes a "statement". This is either true or false.
B (x , y) meaning brothers is a relation, so a state, while B (Peter , John) is the statement that Peter and John are brothers. Only of course, if we know exactly who these two persons are.
Unfortunately, these names mean only properties because there are many Peters and Johns.
So, in math we have to restrict our names from the start to concrete objects or persons.
Suppose that Peter and John are regarded not as such concrete persons, only as properties of having these names. In short as P (x) and J (x). Then, to express not a statement merely a state that two such named persons are brothers, would be quite "logical" as B (P (x) , J (x)).
This is incorrect though! We can not insert states into states, only objects.
The correct way to say this is: P (x) and J (y) and B (x , y).
So, the two variableness comes out clearly and also the "and" connector must be used.
The problem, that everyday names are not unique, is unavoidable by full names either. There are lot of Peter Smiths and John Smiths. In fact, even historical names can be confusing. For example, the statement V (Napoleon) meaning that he was vain, could be misunderstood as merely a property of all the Napoleons. In a mathematical language we define the names and so if we want we can introduce Peter, John, Napoleon as such unique names. Their real meanings, who these are then is irrelevant. The point is that then B (Peter , John) or V (Napoleon) are statements.
If we regard Napoleon merely as a property or in short as N (x), then again V (N (x)) is incorrect and the right way to say it is now: N (x) $\rightarrow$ V (x) .
This is implication, and what we say is that "If x is a Napoleon then x is vain".
So, the missing point before new math, was to use this seemingly overcomplicated method of not to use relations and properties over each other, rather to go through dummy variables.
These dummy variables are the ones that bring out the true nature of the statements.
Not only the internal relationships like "and" , "or" , "if then" but the crucial precise meaning of existence and universality too. Indeed, now with this new variable method, Aristotle's ancient insights can come alive in a new formalism, realized by Frege:
Existence will mean that: "there is an x that . . . ".
Universality will mean that: "for all x . . . ".
The existential and universal quantors will use variables of a state, in fact will stop that variable to be variable. We can not put concrete objects that is names into them any more.
The actual form how we humans went about this elimination of a variable by quantification was not obvious at all. We could have made some marks "at" every occurrence of the variable.
This "at" could be before or after, maybe even under or above. But instead, we repeat the variable at the beginning and precede it with an $\exists$ symbol for existence and $\forall$ for universality.

Thus the accepted ways to tell our previous two claims as statements, are:
$\exists\ x\ \exists\ y\ [\ P(x)$ and $J(y)$ and $B(x,y)\ ]$ and $\forall\ x\ [\ N(x) \rightarrow V(x)\ ]$.
So, while concretization, that is writing names into the variables, that is regarding cases, is the obvious way of getting rid of the variables even formally, quantification is getting rid of them too. Or as we say it, the quantors "bound" these variables. They are not really free variables any more, merely a memory of what they were. This "half baked" bounding idea to get rid of the variability of these variables suggests even more that alternative notations must be used by some aliens.
We might even think, that we could get rid of these variables easily, just as concretizations do, as follows: We put $\exists_i$ and $\forall_i$ in place of all the variables, namely with different i index for different variables. But this would erase the order of the humanoid quantors which is an error.
Indeed, $\forall\ x\ \exists\ y\ \ldots$ has totally different meaning from $\exists\ y\ \forall\ x\ \ldots$
The first claims an existing y satisfying $\ldots$ for all different x, while the second claims a single such y for all x. So this second claim is much stronger. And yet this pre-sequential humanoid quantification is flawed even with this ordered emphasis in mind! Namely, it hides two things in its generally possible $\forall\forall\ldots\exists\exists\ \ldots\ \forall\forall\ldots\exists\exists\ \ldots$ form.
The first is that within a single $\forall$ repeating or $\exists$ repeating group the order is irrelevant.
The much more important second hidden feature is that the universalities do combine in a sense.
Indeed, any statement implies that for all combinations of the universally marked variables, there are some existentially marked variables where the actual quantorless situation is true. So in a sense, the universalities shouldn't been interrupted by the existences. These interruptions merely express on what universalities the existences depend on. Namely, always only on the earlier ones.
So, a much better notation would be to start with all the universalities and then list the existentialities by marking from what universalities they depend with some arrows. This would allow more complex dependencies of the existences. From this we can see that our sequential quantification is simply a short form of the widening dependencies to avoid arrows.
Widening here means that the existences can be put in a sequence that later ones always depend on all those universals that the earlier ones depended on.
Luckily, statements with non widening dependencies can be replaced by more widening ones.
This more general, non sequential quantification concept is not merely a weird idea but can make the proof of the most fundamental Löwenheim Skolem theorem much cleaner and easier too.
By the way, above we were not quite correct because a general quantification sequence can start with a group of existences. These frontal, non depending existences are the only true existences that simply avoid the use of prefixed names in a language.
We also have to admit that maybe in a didactical sense too, the sequential quantification might be a better start. Everyday languages seem to contain a plausibility of sequential quantifications.
Even if this is not true, the ease with which everyday quantifications can be clarified in children by using the sequential quantor notations is amazing.
Don't forget that mathematical descriptions aim at some external realities. No subjectivity, no feelings, no poetry, no humor. But we use this dry section of language in every day situations too.
So an elementary school introduction of this dry language should happen as "grammatics".
I tried out this with my daughter Timea when she was in second grade of elementary school.
"Grammatics" works like a charm and I tried to convince a few people back in Hungary about an introduction of Grammatics in elementary schools. Without success of course!
One day it will be looked back as an absurdity that elementary schools kept on using monkey language, even after the variable clarifications of the quantors $\exists$, $\forall$ were discovered.
I admit that I failed personally too. Timea was smart back then and yet became a lawyer!
A final crucial nomenclature is the "formulas":
Every axiom system has not only its accepted, that is basic names but its basic states, that is basic relations and properties. So all other logically built or "formed" states are called as formulas.
The word "formula" replaced the "logically built states" completely. So the vision that:
States become statements if all variables are concretized or bound by quantors, is now said as:
All formulas become statements. Some don't even use the word statement rather sentence.
And some merely call the statements as closed formulas.

Natural Numbers, Why Strange Models Are Unavoidable

The "encounter" with numbers is the same for a child as for pure math, namely "counting". In its simplest form this means a basic relation of consecutiveness $\lhd$. So behind the different languages and number systems lie the simple facts:

$1 \lhd 2$, $2 \lhd 3$, $3 \lhd 4$, ... Or in short: $1 \lhd 2 \lhd 3 \lhd 4 \lhd$ . . . .

The most obvious thing we can tell is that every number has a next: $\forall x \exists y \ (x \lhd y)$

Observe, how the everyday expression "has" actually hid existence. That's why "grammatics" in elementary school would help to realize these mother tongue tricks early.

The "reverse" of our previous claim is not true. Not every number has a previous. Namely, the number 1 has no previous or more precisely, there is no number previous to 1, or even more precisely it is not true that there is number previous to 1 : $\neg \exists x \ (x \lhd 1)$
More importantly, every other that is not 1 number does have a previous. So, if a number is not equal to 1 then there is previous to it : $\forall x \ [(x \neq 1) \rightarrow \exists y (y \lhd x)]$

These claimed next and previous numbers are unique too! This means that a hypothetical assumption of an "other", is actually the same, that is equal:

$$\forall x \forall y \forall z \ [(x \lhd y \text{ and } x \lhd z) \rightarrow (y = z)]$$

$$\forall x \forall y \forall z \ [(x \lhd y \text{ and } z \lhd y) \rightarrow (x = z)]$$

Of course, now that we have these exact definitions of uniqueness, we can abbreviate unique existences as $\exists!$. Thus our three axioms of consecutiveness are:

$$\forall x \exists! y \ (x \lhd y)$$

$$\neg \exists x \ (x \lhd 1)$$

$$\forall x \ [(x \neq 1) \rightarrow \exists! y \ (y \lhd x)]$$

So, now we might think that the infinite many $1 \lhd 2 \lhd 3 \lhd 4 \lhd$ . . . . name axioms plus these three with variables, will only allow that we actually think about our names in abstract. We can even envision a hypothetical e external or extra number and since it must relate to some other by our first or third axiom, it should follow or precede a normal number.
But since we also claimed uniqueness and no previous to 1 thus this e actually can't exist at all. Unfortunately we were wrong! We assumed only one e but what if a whole bunch of them can exist together? They clearly must relate to each other. In fact each must have a next and each must have a previous. Indeed the 1 exception does not apply to them. So they are an infinite chain both forward and backward, so the negative and emphasized positive numbers and zero can be used:

$$. \ . \ . \ . \lhd -4 \lhd -3 \lhd -2 \lhd -1 \lhd 0 \lhd +1 \lhd +2 \lhd +3 \lhd +4 \lhd \ . \ . \ . \ .$$

In fact, we can have a second, third or arbitrary many new versions of similar sets.
All these can exist next to, or outside our real natural numbers because the three theoretical axioms allow them as irrelevant garbage in our universe. Indeed:
1.) Every element in the universe has a unique next one.
2.) Our 1 has no previous.
3.) Every element in the universe except our 1 has a unique previous.
Realizing that these three "damn" statements are not enough to avoid garbage, is not enough.

You should be angry and try to find better statements. Only then, will you see through your failure why "every" and "there is" claims about consecutiveness can not work..
The simple visual claim : "these are all" is not expressible. It doesn't talk about individuals.
The next best thing to exclude these double infinite "snakes" could be to say that any collection of numbers must have a first. But collections or sets then must be allowed as objects in general.
So we encounter a whole new set of "pain".
Amazingly there was a promising way out without sets!
Before telling this though, I want to emphasize again the predicament from a more general view!

The Isomorphism Principle And Its Failure In Math

There is a lot of talk about "animal intelligence" in popular science programs. They use the word intelligence loosely and interpret experiments suggesting much more in animals than actually is. Communicating and problem solving chimpanzees and dolphins are myths!
Thanks to Set Theory used in a very primitive, finite sense, we can exactly grab the true meaning of human intelligence. It is the simple fact that whatever a child learns, is at once and by itself forming rules in the child's head and is ready to be applied to an infinity of models outside.
Suppose you teach your child to play chess. You have your set of pieces at home inherited from your father and in few days you explain all the rules and play some games. Your child is a beginner. No tricky strategies, just the raw knowledge of the rules. You go out for a walk to relax and in a park you encounter one of those giant chess boards in the ground with huge pieces as wooden statues. And here a miracle will take place right in front of your eyes. Your child will surpass all animals together who ever walked the earth or were trained by over zealous "animal psychologists". Indeed, your child will immediately not only recognize the large pieces but is "ready to play". I don't have to tell more, you get the picture because you yourself have the intelligence. And all this is applicable to any rules that you find yourself or learn from others.
A rat is a very fast "learner". It can explore a maze in minutes and by lucky trials finds the way to the food. It will even remember this for weeks. But what it will never be able to do is memorize the rules as simple sequence of turns as: left, left, right, left, **. . .** and then apply it to a new maze.
Animals live in the full realities of physical appearances. We humans live in models with rules in our heads about these models.
This basic human intelligence should be called the "isomorphism principle".
Iso means same and morph is form so it says that these models are similar. They are different physically but we correspond all the "pieces" as elements, and then their relations remain too.
But, we have to mention a prelude to this isomorphism principle. Way before the child is looking at models he or she is using the mother tongue to express subjectivity, feelings. Doing this he has to follow rules. We know that he does that because he will say "childs" instead of children or similar mistakes. Are the rules of language a preparation for the models? And how much wisdom is in speaking already? The mechanical rules of our mother tongues are so varying and yet they all help in the same direction. To free up the mechanical rules, express thoughts. In fact a parallel but totally opposite line is poetry. It is a reality beyond language and even beyond prose. It is a transcendental residue of feelings and emotions obeying a minimal frame of tongue and turning language against itself. The fact that poetry can arise on the non mother tongue is a deep paradox.
More close to our line is of course the need for the already mentioned "grammatics", to discover Logic behind our mother tongue. This would mean a breakthrough in elementary education.
Two seemingly unrelated questions arise about the isomorphism principle that strangely connect.
First, to degrade reality to mere models is not an obvious advantage. Maybe it would be nice to appreciate the infinity of meaningless details like an animal. Secondly, where are the rules in our head? Are they a template, a super model itself or just the rules exist, the language?
The connection starts with surprisingly and seemingly contradicting my big claim about the isomorphism principle as a final black and white distinguishing feature between animals and humans. Namely, animals possess an elementary version of such principle too. When a lion sees a zebra hundreds of meters away it is a tiny flee on the lion's retina. The meaning of the zebra is an

isomorphism. Of course these analogue isomorhisms are different from the finite rule systems. Strangely, an analogue system requires more nerve cells. So if a deeper analogue versus digital distinction is the base of the isomorphism principle, then the non neurological aspect is even clearer. More towards the connection, the digital isomorphism being a super model or rule system might rely on some weird mixture of analogue and digital modeling. Dreams are exactly such. But animals dream too, so a difference in our dreams is a crucial point.

We have to leave all this behind. In fact, the most important new angle is an opposite of all we said. Indeed our digital or ruled models were meant to be finite like the chess game.
But amazingly we grasp infinity too! Not as easily as these finite isomorphisms though!
Unlike this isomorphism principle that comes alive without any effort, math finds resistance.
We could start here a whole new book about the reason of this, but I have to jump to the point:
The isomorphism principle fails exactly in the most abstract of our worlds, in mathematics.
The point of mathematics is the vision of infinites but infinite models are not isomorphic for our rule systems. The appearing “garbage” outside the naturals, is the nature of infinity.
These weird models are the most mysterious of all things in the universe. Even math can not approach these because plausibility is lost in them.

The Rules Of Logic: Indirect Concretizations And Matrix Contradictions

Beside the acceptance of the strange models as an inconvenient fact, there was a more positive almost status quo, namely that the Hilbert rules are the correct rules of Logic.
I will not tell them now and I will not tell them ever!
This defiance expresses that something is very wrong about them.
And yet, I regard Hilbert as the most positive person in new Math.
An opposite of Gödel the fourth monster giant after Newton, Gauss and Einstein.
But things detach themselves from their creators and can become opposite of what they were originally. So the didactical purification of Hilbert can become a didactical nightmare of today.
And the overcomplicated derivations of Gödel can become the heuristic principles of today.
Of course the original texts are there, so one day a true judgment will be made.
But now my goal is transference of real knowledge that is understanding, not historical judgment.
So we have to go back to the essence of Logic. First I said it is the two quantors but then I said that it was actually the proper use of variables. Well, there is still a vital nuance missing from these variable usages. And maybe this was also part of why variables became clear so reluctantly.
Above in our statements we always used the x , y variables again and again. This is okay because as we explained, these are dead variables, our humanoid quantification skeletons. But now an alternative alien notation system will come in as a more interesting idea because it might be a didactically correct road to avoid the whole Hilbert rule fiasco I am getting to!
The point is that there is a higher level of statements than above as claims about a reality.
Namely, the derivations! We use variables differently when we make whole arguments.
We start as: “Look at this ABC triangle!”. The quantifications here in these arguments are not clearly stated in advance. Instead we stick with our letters all the way through.
Hilbert was re-axiomatizing Geometry when he realized the crucial roles of these variables.
Some call them parameters but I call them “created names” for the following important reason:
What we derive today, somebody may use tomorrow for something else, so if he wants to proceed from our arguments he should not use our letters again rather create new ones.
Hilbert wanted to go oppositely. Avoid these prameters by reducing them to proper variables.
But it is a simple matter of fact that we can go our way, that is turning the parameters into names.
The real truth is that Hilbert knew exactly this way and also the plausibility behind it, just as Euclid new the three plausibilities of parallels. But he wanted to avoid this plausibility to elevate the variables into a new level from mere quantification skeletons to derivational tools.
Had he known what lied behind Gödel’s overcomplicated proof of the Completeness Theorem or how Henkin will simplify it, he would have kept the concept of derivational parameters alive.

Because in spite of the usual wisdom he was not a Formalist in my sense of the word. This sense is very simple: A Formalist is a liar! Either he hides what he sees or says what he doesn't see.
Hilbert said that if someone truly understands something, then he can explain that to the first person on the street. Truer words were never spoken! So lets explain derivations.
Axiom Systems should be regarded merely as starts of Descriptions.
As we claim more and more theorems, our description is growing.
This updated description is what we reason from. So, we don't start from scratch, that is from the axioms, rather use all known theorems. This accumulativeness is common in all science and we accept it as natural. In fact, most outsiders believe that science has an accumulative understanding. That the visions are transferred too. Unfortunately this is false, exactly because of the above mentioned Formalism. Only the irrefutable abstract facts are accumulative. Some Formalists believe that this all right and the visions and understanding comes by digging through the abstract facts. This is false not only morally as a selection of the "successful" students, but also because the deepest problems are exactly tied to understandings. So a future Logic will be didactical!
Until then, we still have to learn and understand this limited Logic of derivations.
Hilbert finalized a variable Logic to do this job. This variableness tells a lot, even without the details. Indeed, the theorems are statements, so contain only dead variables. Thus, deriving with Hilbert's rules means reanimating these dead variables. Taking away the quantors and use formulas. The obvious problem is the conflicting meanings of variables.
Before this, the derivations were informal and variables were used also as parameters having their role only for the particular derivation. To use an old derivation then we should avoid to use these in our own, so there is no conflict. You may say, why would we want to go into the details of already established proofs. But this is unavoidable because to derive new things from old theorems requires a dissection of those. The idea that things just follow from each other like beads on a string is false. Unfortunately, the crucial didactical meaning of the parametric derivations became totally covered up by the Hilbert rules. What's worse, the completeness theorem which is a beautiful consequence of the old natural vision, now is turning up as some magical coincidence.
Luckily, the widening of descriptions as accumulation of theorems, can be exactified and used in two manner. The first is a practical finite "investigative" widening and the second is an infinite "theoretical" widening. The first is how we should define a "raw logic" and the second is how it instantly shows its own completeness.
The raw logic or investigative widening relies on two principles: Indirectness and concretization.
Indirectness is exactly how the non Euclidian geometers and the completeness theorem went.
These of course never reach contradiction, while normally, to derive a C statement we must get a contradiction from adding $\neg\mathrm{C}$ temporarily to our $\mathbf{D}$ description. Indeed, if we arrive at a contradiction, then this means that $\neg\mathrm{C}$ caused the contradiction and so it must be false in every reality of $\mathbf{D}$. This of course means that C is true in all those realities, so it is a consequence of $\mathbf{D}$ and so we can accept it as a theorem and new member of our description.
So, $\mathbf{D}+\neg\mathrm{C}$ leading to contradiction gives $\mathbf{D}+\mathrm{C}$ as new description. A practical generalization of this description concept is that we'll allow not only statements in $\mathbf{D}$ but names too.
All the original names of the axiom system plus new names "created" along the roads to contradictions. In fact, all the reasonings from our temporarily $\mathbf{D}+\neg\mathrm{C}$ description to reach a contradiction, are merely sequences of two simple possible steps of concretizations:
We can pick any S statement from $\mathbf{D}+\neg\mathrm{C}$ and look at the first quantification of S. If it is a universality, say for a variable x then we drop this universality and replace x with any name in the description already and add this new statement to our description. If the first quantifier in S is an existence of x, then drop the quantor again but now create a new name and write this into x and add both the new name and the new statement to our description.
It's obvious that this method is fair, that is it keeps our description true for our reality. In other words if it was true for an intended reality then by these widenings it stays true. Indeed, if something is supposed to be true for all x, then any name must satisfy it, while if something must exist then calling one of those by a new name doesn't do any harm. If that existing object is unique and already has other name, then the equality with that name will be true too. But our widening description could be true for realities with truly wider set of names than the originals.

This suggests already that the potential models that are supposed to prove the completeness of our investigative indirect concretizations as widening descriptions, will be made from the names with the equal ones regarded as single objects. As we explained, the non contradictory or "consistent" statements leading to a model, would prove the completeness of our method. But not all investigative extensions as contradiction searches lead to such model.
The investigative extension can be viewed as method of a real mathematician. He can still have his creative mind. He still chooses the statements from $\mathbf{D} + \neg C$ and chooses the names he uses in universal statements, in any order he wishes. If he has an idea of proof he can choose wisely and reach a contradiction in finite steps.
In the proof toward a model from failing contradictions, we won't allow these choices:
On the other hand, we do allow any extra initial set of names in our initial description, not merely the ones used in the statements. So, we'll have three kind of names. The real ones in our axioms, these ad hoc extra ones and the created ones we have to add for existences later. This will mean that any ad hoc set can be extended to a model. So automatically, we get arbitrary big models even for number theory. But if the ad hoc set is empty then the totality of the names is only a sequence, so as an opposite surprise, even set theory that talks about arbitrary sets has a sequence model.
Instead of the intuitively guided mathematician's choices for particular concretizations from a description, we want to achieve a widening that will guarantee all existences and all universalities to be true among the names. This is a big thing to achieve because an existence, that is an existentially starting statement being true for a name means a new statement being true that may start with again an existence or universality. A universality being true means even more, because the dropped universal quantor again creates a new statement but it is applied for all names in the description in place of the variable that was universal. So we need perfect truths of all shorter statements down to the fully concrete cases of basic states. Which is perfect because then in reverse, building up the statements with quantors, we can see that they are all true back to the axioms. The big question is whether we can achieve such total truth by simple concretizations.
First of all, our concretizations obviously prefer the existential statements because we at once make sure that it becomes true for a name. So, the crucial step is how to satisfy the universalities.
Putting all names into a single universal statement will make sure that it is universally true but it will create a lot of new statements. So we have to create new names again.
As first step, there is a simple strategy to make sure that all the existences and universalities are satisfied alternatingly. Namely, we claim new names for all statements that start with existence. Then we claim all cases for all statements that start with universality. Of course again, every new concretization can have a new start as existence or universality. So we have to repeat this alternating process infinitely. It never gives a perfect set of expanded concretizations but amazingly, combining all these alternating expansions we do get a perfect set.
Indeed, one statement contains only finite many existences and universalities. So all existences will have claimed name case and all universalities will be obeyed by all names. But there is an additional consequence of the finite quantor occurrence. Namely, that all statements run out of quantors. So the total description will have fully concrete cases without quantors too. So, these can define the basic states of all names.
Unfortunately we didn't specify a crucial detail that becomes evident through an actual problem with our basic idea. With no possible contradictions the concrete cases should not have contradictions either. So indeed, we can use these cases as a definition of the relations. Now the assumed meaning of no contradiction is that no finite expansion leads to a direct contradiction. So for the fully concrete cases too, we only know that no finite many of them can be directly contradictory. But why does that mean that the full infinite set is without direct contradiction?
Thinking deeper, we realize that the whole concept of direct contradiction was not specified.
It's quite plausible that this shouldn't be simply a faulty statement. A single statement can be in itself false too, but that's too special. We might say it is instead a relation between two statements namely being opposite C and $\neg C$ pairs. But this is still too narrow. Ten statements can be contradictory too without having such pairs.
So we need a new exactification of our whole Logic and there is a crystal clear road towards this:

Namely, apart from the two quantors, Logic still contains the possible "and"-s, "or"-s, "if then"-s, and so on connectives! Beautifully, these can be combined in a heuristic "situation matrix":
The lines are used in an "and" meaning and can be called as "scenarios" because these lines themselves under each other are used in an "or" sense. The elements of the lines are only allowed to be basic states of the language or negatives of such. So, the first drastic shift of this vision is that all negativities are pushed into basic states. The second shift is that implication, this classically so overemphasized logical connector is missing completely. This is a good thing!
Implication is a confusing thing. The $C \to D$ form is not some deep cause and effect relation!
It claims that if the C condition is true then D is too. So if C is false then we claim nothing.
$C \to D$ is automatically true. Thus in effect what we claim is C being false or D being true.
The only true meaning of $C \to D$ is: $\neg C$ or D. So $C \to D$ should be an abbreviation of this.
The pushing of the negatives deeper and deeper is possible because the negatives of "and"-s and "or"-s are easy to express with each other. Indeed a C and D being false simply means C or D being false, while a C or D being false means that both C and D are false. So:
$\neg\,(C \text{ and } D) = \neg C \text{ or } \neg D$ $\qquad\qquad$ $\neg\,(C \text{ or } D) = \neg C \text{ and } \neg D$
The naturalness of these are the reason that the final matrix form is the winner.
The lines are the "and"-s and they themselves connect by "or"-s.
To see how beautiful and revealing this system is, lets formulate the classical recognition of Euclid that there are infinite many primes. As language, we need the naturals as names plus equality, smaller bigger and multiplication as basic relations: $x = y$, $x < y$, $x \bullet y = z$.
The composites are numbers that can be written as products without using the trivial 1 number.
So y is composite if $\exists u \exists v \ (u \neq 1 \text{ and } v \neq 1 \text{ and } u \bullet v = y)$
This means that 1 itself is not composite but it's merely because it's too simple. The other non 1 numbers that are not composites are the primes: 2 , 3 , 5 , 7 , 11 , 13 , 17 , 19 , . . .
The infinity of these means that they must become arbitrary big, and in reverse if they are arbitrary big then there must be infinite many, so our claim can be said by:
For every number there is a bigger non composite. Or in formal language:

$\forall x \exists y \ [\, x < y \text{ and } y \text{ is not composite} \,] =$
$\forall x \exists y \ [\, x < y \text{ and } \neg \exists u \exists v \ (u \neq 1 \text{ and } v \neq 1 \text{ and } u \bullet v = y)\,]$

So we encounter an other pushing in of the negatives, namely through the quantors.
This is very plausible too, in fact Aristotle already collected them as rules.
"Not every" means that "there is some not" which of course is not "there is no" or "none".
In reverse though, "not there is" means that every is a not so indeed "none" is.
In our case the non existence of u and v means that for all u and v the composition fails:

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

Of course, here it's easy to push in the negative as we explained by changing the "and"-s to "or":

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

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

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

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

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

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

This finally in a matrix notation is:

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

Now we can return to our dilemma about the exact meaning of contradiction.
After the quantification concretizations, we achieve an infinite set of concrete matrixes.
This means having original or created names in them in place of all variables.
Each matrix is a claim of its lines as scenarios that is alternative “or”-s. So the lines are actually choices and so a finite set of matrixes should be contradicting if we can not chose lines from them without those being contradicting. But what is a contradicting set of lines then? That’s easier!
Each line is “and”-s of basic states or negatives and these are already concretized so contradicting lines means the appearance of exactly opposite basic states.
Like $(a < b)$ in one line and $\neg (a < b)$ in an other. These a , b are basic or created names.
It’s easy to spot such contradiction but don’t forget the possible lines! One line selection being contradictory is not enough. To see that some matrixes are contradictory you have to try all possible line choices from them and show that they all fail.
It’s also clear now that for an infinite set of concrete matrixes if one can chose lines from each so that no contradiction is among them then they instantly can be used to define the model:
We simply regard the basic states or negatives listed in the lines as a definition of the basis states.
If some cases are not specified either as true or false in them, we can choose as we wish.
So a freedom exists and the matrixes themselves had some freedom in the concretizations already.
But there is a third one too! Lets not forget why we went deeper at all! The consistency of a system means no contradiction in any finite set of matrixes but the model creation meant line choices from all the infinite many. Does this always follow from the first? Yes!
A Set Theoretical argument using the earlier mentioned Wellordering Theorem not only proves this but gives a heuristic strategy how to choose the lines and thus how to create a model.
Since the Wellordering of the matrixes is not deterministic, we have the third arbitrariness mentioned above. By the way, this principle that guaranties the infinite consistent set of matrixes that defines a model, is not depending on all details of the matrixes.
Only three important points matter that are even clearer with a new meaning:
Every matrix has finite many lines. Like buckets that each could contain finite many words.
The contradictions can be any predetermined set of sets each containing finite many lines.
That is: A set of sets is given, each containing finite many words that are taboo together.
Finally, we simply know as a fact that:
All finite subsets of matrixes have choosable lines avoiding these predetermined contradictions.
That is: From finite many buckets we can always choose words that avoid all taboos.
If these three finiteness conditions are true then:
There is an infinite choice of lines from the matrixes that avoids contradiction. That is:
We can choose words from all buckets that the infinite text will not contain taboos either.

A Concrete Concretization Sequence For A Single Statement

Above I just sketched the whole essence of logical derivations through three stages:
1. Negating the statement that we think should be true and adding the negative to our description.
2. Quantor concretizations using new names for existences or old names for universalities.
3. Finding matrix contradictions among concretized matrixes.

The justification for the correctness of this method is when we don’t reach contradiction, rather create infinite many statements having also infinite many concrete that is quantorless ones among them without contradictions and then these create a model.
But these create a model only if the created full set of statements is a perfectly step by step concretized set. That is, every existential statement has a cased or named version and every universal statement has all its named versions inside.
To achieve such perfect set of statements, we can use the heuristic alternative existential universal concretizations. These are never perfect because they create new statements but their combined total is actually perfect.
The second stage also requires a heuristic method, namely we have to demonstrate that any finitely non contradictory set of matrixes always allows an infinite set of line choices. I said that this is a basic set theoretical fact that can also be regarded in general as taboo avoidance.
So, we can leave the visualization of this second part to later, inside set theory.
On the other hand the first part, that is the alternative concretization, in spite of being so simple, is quite surprising and so should be shown more concretely.
So, I will demonstrate it with a single starting statement: $\forall \; \exists \; \exists \; \forall \; \ldots \; 1 \; \ldots$
I omitted the variables but indicated that the single 1 name is used in the statement too.
We can continue to use the naturals as created names. Since the statement starts with universality, first we have to claim all universalities for the already existing names which is just 1 now.
Then we can introduce 2 and 3 as new names because we have two existences. And so on.
The concretized versions will be listed with putting the names under the corresponding quantors:

$\forall$	$\exists$	$\exists$	$\forall$	
1	$\exists$	$\exists$	$\forall$	$\forall$-namings using 1
1	2	$\exists$	$\forall$	$\exists$-namings = making new names from 2 , 3 , . . .
1	2	3	$\forall$	
2	$\exists$	$\exists$	$\forall$	
3	$\exists$	$\exists$	$\forall$	
1	2	3	1	$\forall$-namings using 2 , 3 in old ones and 1 , 2 , 3 in new ones.
1	2	3	2	
1	2	3	3	
2	4	$\exists$	$\forall$	
3	5	$\exists$	$\forall$	$\exists$-namings = making new names from 4 , 5 , . . .
2	4	6	$\forall$	
3	5	7	$\forall$	

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

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

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

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

Concretization Sequences For The Naturals

The previous example was very concrete but fairly useless.
Though finite many axioms can always be replaced by a single statement, the usable axiom systems are always infinite.
In fact, already our name axioms $1 \lhd 2 \lhd 3 \lhd 4 \lhd$ . . . . . are infinite.
Soon we'll introduce an other infinity of axioms.
Those formalisms that use the consecutive number of an x as an s (x) successor function, regard only 1 or 0 as names. So, we can avoid the infinity of the name axioms.
On the other hand, this s is then has to be used in the matrix logic repeatedly. To allow functions in general, makes things even less transparent. So I rather take the burden of infinites.
The strange models of the naturals are also better "visible" this way. Of course, our previous example then must be modified at once, because the naturals are already used up. So we must introduce a second sequence of names like $\underline{1}$, $\underline{2}$, $\underline{3}$, . . .
As we saw, the weird naturals are always double infinite snakes but our model method can create those from this single sequence. This sounds quite unbelievable from an other angle too.
Indeed, the axioms are themselves infinite or as we said they will be even double infinite.
To avoid this double infinity of course is quite easy by combing them together like the even and odd numbers to form a single sequence of axioms. But this infinity itself means that for the existential stages we need already infinite many new name sequences if we go by the way we described it, that is create new names simultaneously for all existences.
A much trickier way is to introduce the new names in a single sequence by going ahead in our statements more and more. So, we envision the statements as a sequence from left to right and we'll make the concretizations under them but always going ahead just one and return to universalities for all earlier ones. This idea is called "dovetailing". We'll use it later too.
The end result is that we get a model with this clear distinction of the real and fake naturals as two sequences. The fake sequence of course has to become sets of double infinite sets and if we have a $<$ relation then they all have to line up. We can even tell some facts about how this happens. But the sad fact is that we have no clue what weird truths could come about in our full model caused by the fake naturals. The two heuristic methods of the alternative concretizations and the set theoretical line selections reveal nothing about the final model. But now with these remarks at least you can see why we don't see anything.

Special Features In Matrixes

An important but yet unmentioned detail about contradictions is the following:
If only the perfectly contradicting cases as $(a < b)$ and $\neg (a < b)$ would give contradictions then contradicting fully universal assumptions or axioms could be instantly recognized.
Indeed, universality allows all name replacements and so we could simply write a single letter in all variables. So, a tell tale sign of contradictions would be opposite basic states like $(x < y)$ and $\neg (u < v)$ because it gives $(a < a)$ and $\neg (a < a)$ with using the single name a.
But this is not so because we have equality! This particular relation can be contradictory in itself.
Namely $\neg(a = a)$ is a contradictory concretization of $\neg(x = y)$. This then stops the trivial collapse of variables into same names and thus makes things more complicated! This of course still keeps the finiteness of contradictions, so the big picture as taboo avoidance remains.
Universal axioms still mean a simplicity because all concretizations can be used.
Thus we don't really need names as concretizations. Combining of variables themselves could be used. So here an instant variable logic principle is identical with the heuristic name concretization.
Hilbert and Ackermann deduced this principle backwards from their variable rules and this was known by Gödel. He went from this particular case to the general model construction.
Having infinite many universal axioms, it is still not obvious to defy a contradictory finite set of matrixes. If however we have only finite many universal axioms, then we can first combine them into a single matrix and then we only have to try out all possible combinations of this.

So we can tell mechanically if contradiction is avoidable.
This seems to contradict that there are finite universal number theories that as we'll see, shouldn't allow such decision method. The reason is that they don't use consecutiveness as a relation.
With consecutiveness, we don't have only universal axioms. Namely the first and third axioms of consecutiveness must claim existence. Plus our system is not finite either because we have the infinite many name axioms: $1 \lhd 2$, $2 \lhd 3$, $3 \lhd 4$, ...
The fully universal finite systems "cheat" twice. First, use an s successor function instead of consecutiveness. This assumes the existence of s (x) automatically and can avoid the infinite many name axioms too. You need only one name, 1 and the other numbers are repeated applications of s : 1 , s (1) , s (s (1) , . . Worse yet, is to use 0 instead of 1. And finally: Using < allows to hide the existence of previous to non 1 or 0 numbers as universality too.

Returning from universality to general systems, the main difference is that here the quantification orders determine the possible name concretizations and thus also the allowed name combinations in the matrixes. So the question of a contradictory finite set is even more complicated.
Only this is the point where a functional approach like the s successor could be introduced to get a clearer vision. Indeed, these functions can turn all systems into universal formally.
Back to our main subject. With this heuristic identity of the Logic of concretizations and the creation of weird models, we see even better how the isomorphism principle is failing in math.
The fundamental question is this: Are these weird models meant to be? Not just in general but in their details. With their particular individualities. This is still a mystery that nobody knows!
The easier question is this: What was the mentioned false hope for perfect axiom systems?
The garbage extra objects outside our natural numbers are an unavoidable nuisance and mystery, but what if they are irrelevant in a sense. Namely, just like our heuristic isomorphism principle allows us to think in rules instead of models, maybe the perfect rules of infinites though allow weird models, also allow to derive all truths. So all statements are decidable from axioms. Then these other models are merely phantoms because they all have the same truths. For this to work, we would need a heuristic method of finding new useful relations among the natural numbers and a heuristic method to claim axioms about these. In fact, as I hinted it already in the previous section, we probably would need infinite many more and more complicated basic states and infinite many more and more complicated axioms about each of them. This double infinite system does exist, but before I reveal this, I make a detour in the next four sections about Proof Theory. This is the field of direct derivation systems, used instead of the indirect concretizations.

Beyond Logic

The strange models as side result of our heuristic concretization method is not the only mystery.
A much simpler one is real mathematics. Thinking in our intended models!
So it's crucial at least to see, where we avoided this deep jungle and in a sense lied to claim that our concretizations are the true logic. And we can tell this point clearly! Namely:
We changed our statements into quantified matrixes. We never think this way!
Our example was the infinity of primes: For every number there is a bigger prime.
This English version of our claim says only the two "main" quantors. The rest are hidden in the meaning of the prime. And in a sense this is how we think too. We only think about these two quantors directly and the ones inside the concept of primes are internalized as plausibilities.
This at once suggests that for a better Logic we should allow the quantifications everywhere.
This of course means that the "and"-s and "or"-s are not finalized in the matrix either.
Then a crucial question is whether we should also allow to free the negations from the prison of the matrix or not. Should we allow them everywhere? For now, lets say we do this too.
So, total freedom of formulas combined, that only turns out to be a statement because the variables are all bound if we look closer. Doesn't this mean that indeed the formulas are the bigger field in which mathematical thinking could be found? Maybe this is so. But the variable logic of Hilbert was definitely not proceeding by the full mathematical thinking.

It was merely a restricted imitation of our indirect concretizations done directly and for variables. Its sufficiency seemed as a mere luck and not heuristic at all. In other words, the completeness of it is not plausible and hard to prove. And these Hilbert rules did become replaced by Gentzen. Strangely, not with the direct intention of getting a simple system that is trivially complete.
The intention of Gentzen was dual. To be more close to real mathematical reasonings and also to be able to show that the system is consistent. This second obsession over rid the first and a third new aspect appeared by itself. Namely, that there are alternative avenues in derivations.
Some rules are correct, even frequent in real mathematics and yet avoidable in some sense.
So avoiding these, we get a minimal system. Most amazingly, the rules of this minimal system are all "finitely reversible". To see what this means is very helpful to understand what's coming.
Normally, a rule system means that we can derive the theorems by smart choices of the rules.
We can also understand that since we only have finite many rules, we can apply them all as trials and so we can generate all theorems mechanically. Unfortunately this is not true because our axioms that we start from are usually infinite many. But lets say we can generate them too and so we can truly generate all theorems by a computer. Then we can also tell the computer to watch for the statement that we suspect to be true. If it comes up the computer will tell us. Voila, computerized mathematics. And this is true, in fact, the idea can be extended and we can assume that we might be wrong and so as a "safety" precaution we tell the computer to watch not just for our statement but for its negated version too. Then we can realize that this safety was not really good because the computer can still run for ever. Indeed, it could be that neither our suspected claim nor its negation is theorem. The claim is undecidable. This is exactly where we heading but right now our point is much more practical. Namely, that with a good Logic, the computerization is not as stupid as we pretended. Indeed, what if the rules are such that regarding any formula there are only finite many possibilities of what possible assumptions could lead to the formula by the rules. Namely, because those possible assumptions are all recoverable from the formula.
A formula can only hide finite many such recoverable possible assumptions. So we can reverse our rule system with these finite choices and trace back if indeed a derivation was possible.
Observe that these finite back choices will still lead to a potentially infinite back search and even will branch to different directions if we reverse rules using two assumptions. But at least we'll get a single infinite search. Most importantly, we start with the target and so this should be faster than generate theorems blindly and compare each with our target.
Observe that indirect concretizations also started with the target but negated and added to the basket of our axioms as starting description. Then the derivation of a contradiction came by concretizations applied to blindly chosen statements from the basket. If we start with an empty basket that is no description, then the contradiction means an impossible statement and so the original negated claim was a logical necessity that is true in all models. But even these logically true statements will be hard to establish because the concretizations create new statements.
Of course our proof of the completeness theorem showed that we can concretize systematically.
This is exactly what the beautiful minimal rule system achieves here directly and that's why its completeness will be so easy to see from its reversal. To be honest, this reversible minimal system works only for logical truths. They can be used backwards successfully only from a single assumed logically true statement. This at once raises the question "back to what success?".
After all there are no description axioms if the target is logically true. The back end, that is the root of the derivation for these logical truths are logical axioms, trivially true formulas. These correspond to the contradictions in the indirect method. So as we see, this mew derivation logic is very much a mirror image of what we already have but there are big advantages here. Namely, in spite of that the minimal system only works for logically true statements, we can tell that even when we do have non logical axioms, the extra rule is only unavoidable for these. Then it can be avoided or used as logical trick too. So, we do get insight into the derivations which was Gentzen's original goal towards proving consistency.
The most striking difference from Hilbert rules is the use of "and" and "or" instead of implication. But in Gentzen's system the abstraction covered this and implication crawled back at once. Only William Tait discovered three decades later the truly minimal system that I will show in the next section. Maybe he was too humble but he failed to stress the importance of his own discovery.

Proof theorists don't care about the didactics of what they already know because they look for more abstractions. Didactical approach is a science on its own not existing today in our world!
I claim that this didactical science is the foundation of the universe. Our world is a black hole!
So, are there any proof theorists who at least see that something is wrong?
I was only impressed by some admissions of a leading proof theorist Jean Yves Girard.
But he is too smart for his own good and the didactical angle is missing completely from him.

The Root Of Proof Theory

In our heuristic concretizations, we regard the $A_1, A_2, ..$ axioms as an initial **D** description.
To prove a C theorem means to form $\neg C$, add it temporarily to **D** and try to derive a contradiction from $\mathbf{D} + \neg C$. If we succeed then we can add the true C to **D** and use $\mathbf{D} + C$ as our wider description. We can even use this for infinite many $C_1, C_2, ..$ theorems. Then a new N claim can again be negated but now we add it to the wider $A_1, A_2, .. + C_1, C_2, ..$ and try to create contradiction here. Of course, only finite many $A_1, A_2, ..., A_m, C_1, C_2, ..., C_n$ can be used in achieving a contradiction. These used C-s had their own earlier finite many A-s used in achieving the contradictions from their $\neg C$ negatives. So, we should be able to derive a contradiction from all these A-s and $\neg N$. But this has no formal method. A major negativity compared to direct derivations. Indeed, from a direct derivation of N, we can instantly eliminate any used $C_1, C_2, ..., C_n$ earlier proved theorems and obtain one from $A_1, A_2, ..$
This might sound as a dead sentence against our heuristic indirect method or as an instant success for direct derivation methods. But we are wrong. Direct methods can not succeed in clearing up all indirect arguments. Namely, we can use $C_1, C_2, ..., C_n$ intermediate statements for an N claim where we don't prove these C-s as theorems. We don't even assume them as true.
If all we know is that at least one of them must be true, that is their "or" is true then we can add them plus $\neg N$, one by one to our description. If each of these $\mathbf{D} + C_i + \neg N$ leads to contradiction, then N must be true in all realities of **D**. Indeed, in all realities of **D** one of the C-s is true and thus by the derived contradiction $\neg N$ is false. Then of course, by the completeness of our method, there must be a contradiction derivation from $\mathbf{D} + \neg N$ too. Here again we have a situation where our method tells nothing concrete about this derivation. But now the direct derivations are just as useless! Indeed, derivations for the individual C-s don't even exist now.
Of course their "or" is derivable but a derivation of C_1 or C_2 or .. or C_n requires A-s nothing to do with the ones that are needed to be added to the $C_i + \neg N$ -s to get contradictions.
Or alternatively, needed to derive the $C_i \rightarrow N = \neg C_i$ or N implications.
This is evident from the typical use of this logic as "derivation from cases".
If for example we want to claim something about all living creatures then we can show that being men, women, child, animal or plant each imply the claim and so it is valid because all living creatures are one of these. The proof that all creatures are these, is irrelevant to our claim.
The most obvious of this is if we have only two cases, that is C-s , namely C and $\neg C$ because then their "or" is trivially true. In fact, if the general "or" form is true then we can approach it by repeated C or $\neg C$ dual cases as C_1 or (C_2 or .. or C_n) and so on.
These two, the general C_1 or C_2 or .. or C_n form and the trivial C or $\neg C$ special situation are the roots of Gentzen's grand result that replaced the Hilbert rules with proof theory.
The general "or" form is the root of a better system and the special C or $\neg C$ is the root of the mentioned avoidable extra rule. This better system was later simplified to the minimal by Tait.
The aim is a direct derivation system using formulas, just as the Hilbert rules did but these new rules will use the "and" and "or" as opposed to the old implication.
In fact, the real importance is the "or"-s, expressed in both ends of the derivations. Namely:
We always derive C_1 or C_2 or .. or C_n formulas and we start from the same kind as logical axioms if among the members, two C_i and C_j are actually B , $\neg B$ pairs of basic states.

Indeed, such pair's "or" is trivially true and this makes the whole or-formula true too.
It might seem weird to start from this unbelievable wide set of axioms that practically contains everything already. A more humble form of the logical axioms would be to claim just the B or $\neg$ B formulas. Then we could add as the most trivial rule that any or-formula can be "widened" or as they call it "weakened" by adding any new members. The result is the same and this widening is the crucial "ad hoc" heart of the whole direct derivational method. With the wide axioms at least we express that we could make all the ad hoc choices at the bottom of the formulas.
If we have equality then this has its obvious axioms too and we may have our non logical axioms describing a reality. With non logical axioms present an extra rule is unavoidable, though can be limited to be used for these non logical axioms. Now we assume pure logic, in fact for simplicity we'll disregard equality. There will be only three rules, two for the quantors and one for "and".
The best way to visualize these inference rules is luckily the notation itself, accepted by everybody. Above a simple horizontal line we write the assumption or the two assumptions with empty space between them. Under the line we write the inferred consequence.
In Tait Logic we use the inferences to "introduce" the two quantors and "and".
So, in the assumed formula above the line some variables will be regarded and under the line we infer new quantors for those variables. For "and" we assume two formulas above with C and D sub formulas and introduce their "and" that is (C and D) under the line in the inferred formula.
Existence can be introduced for any variable of any member in an or-formula, while universality only if that variable is not appearing in other members of the formula. The existence introduction is plausible because the variable already meant an assumption of existence. The universality on the other hand seems a big jump from a single case. But now these variables are always disguising universality too and so only the relativity to the actual or-formula matters. Not appearing elsewhere means that it is unrestricted inside the or-formula so if the whole formula is regarded with universal variables then this single member is universal already.
The introduction of (C and D) requires the assumption of two or-formulas that are the same except in one member. One has there C , the other D. Then we can infer the new or-formula that is again the same except in place of the C , D it contains (C and D) as member.
As we see, we talk about or-formulas as sets with members so their order is irrelevant. But an other thing that is irrelevant in sets, is repetition. And indeed, two same members of an or-formula can be combined too. So we have these hidden permutation and contraction rules too.
And yet we'll regard these or-formulas very specific in their order. Indeed, the mentioned finite reversals can be made unique if we regard the reversals always on the first possible member.
This means being a non basic state or its negation, that is being a quantored or "and" formula and thus one of the three rules can be reversed for this first member.
The validities of our rules for realities are obvious and their sufficiency, that is the completeness of our system goes similarly as for the matrix method. We will show that if a statement is not derivable then there is a reality where it is false. Now we don't have new names but again the sequence $\underline{1}$, $\underline{2}$, . . will be used as new variables in strict order to create this reality.
Namely, the three rules are used backwards for the mentioned first possible member as follows:
The universal quantors are replaced by a first new variable while the existential ones are also replaced but not with new, rather a "first" old. This "first" means a first that was not yet used for this exact existence. This makes only sense because the second most important difference from universal reversal is that here we keep the existence and put it to the end of the or-formula.
Amazingly, these kept existences will crawl forward from the end as we use up the earlier members and will be encountered again and again as first in line to be reversed. So, we have widening or formulas as reversals. Even more importantly, this trick also hides the dove tailing and ensures that in the end all variables are used as replacements into the existential formulas. The derivational direction will create repeats of the existential formulas but this is simplified by the hidden rule of contraction.
These two variable replacements seem quite the opposite of the quantor meanings but remember that we now strive for falsity of our original statement. And indeed, one false case of a formula implies falsity of the universality, while only all false cases imply the falsity of existence.
The "and" rule reversals are obvious as branchings to use both members of an "and" separately.

Of course, with the encountering of new quantors, this means branchings of widening or-formulas. The non derivability of our initial formula means that at least one path exists without encountering logical axiom, that is containing opposite basic states. Thus, the final definition of our reality is quite simple as those variable combinations of the basic states that don't appear in this path.
So, this simple system is complete, it can derive everything that should be derivable.
Now comes the interesting part! The "and" introduction rule could be used in a new way.
Indeed, if the C , D members from the otherwise same or-formulas are accidentally each other's negatives, that is $D = \neg C$, then C and $\neg C$ = false and so we could leave out this from the inferred or-formula. So a new rule could be to infer this simpler or-formula.
Its name should be "cut rule" because it cuts out this C member of an or-formula if the negated $\neg C$ membered or-formula is derivable too.
The first bad thing about this rule is that we have to establish this negativity. Indeed, we only allowed $\neg$ to appear for basic states. So here in the cut rule the $D = \neg C$ condition is actually the process of moving the negation down to the bottom in C and then seeing that it becomes D.
The second much worse feature is what we already hinted, that unlike the other rules, this is not finitely reversible. Indeed, we can not tell from the inferred formula what C was cut out.
So, a reversal would have to try to put back every possible formula which is infinite many trials.
The completeness argument above proves that this cut rule is not necessary. That is, any proof that uses it can be replaced by a proof without it. And indeed, many proofs use the negativity assumption or more generally cases. We may even say sadly that, what's the use of an avoidability of cut if we don't know how to avoid it. Luckily, Gentzen not only proved that the cut rule can be eliminated rather his "cut elimination" theorem showed how it can be eliminated.
Our completeness argument relied on having only logical axioms. And indeed, if the non logical ones contain "and"-s then it's easy to see that we need the cut rule to break these down.
Luckily, it's also easy to see that we only need to cut such internal formulas inside the non logical axioms while any truly logical use to cut new formulas can be still eliminated.

A Concrete Reversal Of A Derivable Or-formula

$\forall x \exists y \forall z (B (x , y) \text{ or} \neg B (x , z))$ is a logical truth regardless of the B basic relation.
It is derivable by our rules from a logical axiom and our reversal technique shows how.
Remember that $\underline{1} , \underline{2} , \underline{3} , \ . \ . \ .$ are new variables here and start from the bottom:

$$B(\underline{1},\underline{1}) \text{ or } \overbrace{\neg B(\underline{1},\underline{2}) \text{ or } B(\underline{1},\underline{2})}^{\text{Truth}} \text{ or} \neg B(\underline{1},\underline{3}) \text{ or } \exists y \forall z (B(\underline{1},y) \text{ or} \neg B(\underline{1},z))$$

$$\overline{B(\underline{1},\underline{1}) \text{ or } \neg B(\underline{1},\underline{2}) \text{ or} \forall z (B(\underline{1},\underline{2}) \text{ or} \neg B(\underline{1},z)) \text{ or } \exists y \forall z (B(\underline{1},y) \text{ or} \neg B(\underline{1},z))}$$

$$\overline{B(\underline{1},\underline{1}) \text{ or } \neg B(\underline{1},\underline{2}) \qquad \text{or} \qquad \exists y \forall z (B(\underline{1},y) \text{ or} \neg B(\underline{1},z))}$$

$$\overline{\forall z (B(\underline{1},\underline{1}) \text{ or } \neg B(\underline{1},z)) \qquad \text{or} \qquad \exists y \forall z (B(\underline{1},y) \text{ or} \neg B(\underline{1},z))}$$

$$\overline{\exists y \forall z (B(\underline{1},y) \text{ or} \neg B(\underline{1},z))}$$

$$\overline{\forall x \exists y \forall z (B(x,y) \text{ or} \neg B(x,z))}$$

Proof Theory Beyond

The necessity of cut rule for non logical axioms may seem as a minor problem because we only use finite many such axioms in a proof. But this is not so because we don't know which finite many we need from the infinite many. For Gentzen's original obsession of showing consistency, this is even more evident. The backwards derivation from a logically true statement to logical axioms show that this logic is non contradictory. Indeed, the basic pieces must be in a formula that will have opposing but identical variable variations. But a simple falsity can not contain this.
The non logical axioms can bring in an inconsistency very easily but can we detect this easily too? This was Gentzen's question for number theory. His plan was to show that the infinite many non logical axioms here allow a similar backwards widening, so the derivable theorems again already contain the used assumptions. But this is not so. The cut rule cuts deeper among these axioms.
The simple falsity could be derived from arbitrary long non logical axioms.
In the end, he still achieved an amazing consistency proof of number theory using set theory.
Gentzen himself started with natural systems that use all logical operations unrestricted.
In opposition to the heuristic Tait system, this means that negation is allowed everywhere not just at the bottom. Indeed, if we have negation then implication is definable simply as $\neg$ C or D.
Trying to stay "natural" was the reason that he missed the Tait system and instead his system became much more complicated. In this, the three fundamental rules are not enough.
We might think that we need only an extra fourth rule for $\neg$ introduction but this is not so.
We can not trace back the negations from the derived formulas without tracing back the back tracing itself which means that we have to introduce a new symbol for this too. This seems very similar to implication but it is a meta symbol. In fact, if we allow negation then of course we at once have real implication too that we might want to abbreviate and so we need two symbols.
In the original Gentzen system both above and under the line we have the new meta symbol of derivation. These full claims above and under he called sequents.
Some books use $\supset$ for old implication and $\rightarrow$ for sequentation. Some keep the arrow for implication and introduce a new symbol like $\Rightarrow$ for sequents or use the $\vdash$ derivability symbol.
The first important point is that the assumption side of sequents will be not or-formulas rather and-formulas. This might sound insane first because this way we are not proceeding to same kind of formulas and so we could not use any earlier results. But this is not so because we are now proceeding to obtain not formulas rather these sequents themselves. To obtain formulas we simply regard the sequents with empty assumption.
To worsen the whole situation, in order to simplify the rules, we actually don't specify formally the two sides of the sequents. So, we merely regard sequences of formulas both before and after the sequent symbol. Since the orders are not important, we sometimes regard these sequences as mere sets but with allowed repetitions, also called as multi sets. The "and" meaning of the assumption side and the "or" meaning of the consequence is thus hidden and only comes out of the rules themselves.
The "introduction" of the quantors, negation and the connectors "and" , "or", "implication", now use a sequent without these above and then making them appear in the sequent underneath. Just like in the simple Tait system. But now this means two rules for each, namely to introduce it in an assumption or in a consequence. To simplify the names, only this second is called "introduction", while the first is called "elimination" because in an inner sense these logical symbols are disappearing from the conclusion. This means already twelve rules. Plus we need the permutation, contraction and weakening. Finally, the cut rule is cutting out a member from an and-formula, that is sequent condition if the members without the formula were consequences. Which is of course an or-formula of them. But the meaning is evident because if those other conditions as "or" is true then one of them must be true and so they together imply the same as with the cut out formula.
The crucial point is that here again, only this cut rule looses any sub formula. The other fifteen rules are all finitely reversible. The "final destinations" of these reversals are again the logical axioms which are now the sequents with identical condition and consequence, meaning of course that "and"-s of some formulas always imply the "or"-s of them.

After Gentzen discovered this sequent calculus, one would have assumed that the old "natural" systems die out. Indeed, they all lack the beautiful separation of a reversible set of rules that are already sufficient plus the single extra optional heavy gun rule. Yet, the worst old "natural" rules of Hilbert came back from the dead, because an amazing similarity was observed with Effectivity formalizations. This abstract line is still under continuous investigation but has nothing to do with the other "motivation" of still using the old rules. Namely, being more "natural". Which, is a lie!
These "Logic Educators" are the new versions of the medieval "logists" who used Aristotle's syllogisms as actual wisdoms and even combined it with metaphysical blubberings about God.
This word "metaphysical" originally meant philosophical but already Newton used it with this negative meaning when he put as motto to his book "Physics Beware Of Metaphysics". Hegel responded smartly by saying that Newton actually warned: "Physics Beware Of Thinking". But even though he was right in some respect, he missed the point, because he never understood Newton. He never understood what physics is. This is evident from his claim about the "logical" number of planets. There is no logic of course, because it is accidental. I wonder how he would twist his words if he knew what today everybody gets on a silver platter, that there are other solar systems with totally different number of planets. In spite of this, Hegel was one of the smartest persons ever and the clearest idealist. In a sense, his ignorance of the new dialectic hiding behind physics was that allowed him to remain an idealist and thus conserve a deeper truth against the spreading materialism induced by science. Hegel also made a scathing attack against the false metaphysical blubberings of the medieval "logists".
Earlier I mentioned how desperate I was to introduce Grammatics in elementary schools.
As much as the ability to form statements with variable quantifications and logical connectors is vital, the obsession of deriving them systematically is false. Our reasoning will not get better by these idiotic rule systems. Learn real mathematics, especially by solving word problems!
Such "algorithmic abilities" are the keys toward didactics. Literacy and numeracy is the first but shouldn't be the last "must". I call these "musts" the "roads". These are not optional. They can not be graded or marked. You simply possess it or not. The "gardens" are the optional subjects where kids can wonder into if want. Finally the "maps" educate them externally about the big picture.
Educational "gurus" without concrete didactics are well intentioned and usually promote openness and creativity. But missing the importance of didactics, this totally impersonal and effective root of learning is lethal. There is a very simple fact that everybody who tutored hundreds of students like I did, must have realized. Praising, consoling or trying to be nice in any emotional way towards the students is immaterial. There is only one sure way of motivation! Success! Being able to possess the "roads" that everybody can, is indeed the "road" subjectively too!
But the earlier mentioned logic educators are much worse than the well intentioned teachers without the didactical knowledge. To spread useless algorithms as clarifications is direct poison.
The subjective algorithmic ability of quantifying everyday sentences is a "must" because these quantized forms are reflecting the absolute. Not only all mathematical but even everyday claims have this spine. In mathematics, we don't think in these forms either but we are able to translate all precise claims into these quantized forms. Then a proof can be still very messy and didactically wrong but the correctness of a proof is verifiable. We don't have yet a didactical logic and so our logics are all merely logics of verifying derivations. This is non absolute but exact, because we go back to axioms. In everyday sentences we can go into circles because we have no basic states. But the point of quantification still remains because proof is not the point, rather the exactification of our claims. We simply can not avoid this stage even if we think about God, soul, love, jealousy anything at all. Of course, in these philosophical ponderings the quantifications are not important and so the ability to exactly quantify is irrelevant. But the point of elementary school Grammatics is not introduction to philosophy rather introduction to mathematics. Not from merely numbers but from language. The quantifications are all there, hiding in everyday language. What is missing is the variables. Bringing these in is the crucial step towards math. Then combining these with numbers is the equations and the instant return to reality, everyday language, is the word problems. So Grammatics and Word Problems are becoming a full circle, a total entry into math. Everybody who can translate everyday sentences into quantized forms with variables and can

solve word problems through equations, is already a mathematician. Because being a mathematician is not knowing a lot of mathematics rather being able to think mathematically.
So, again we just emphasized more elaborately that variables are the essence of math.
Remember that this was the missing point by Aristotle too and realized only by Frege.
Aristotle did realize the importance of the two quantors "every" and "some" but used these as operations on the full claims. This leads to that individuals and properties or relations are not clearly separated. Now, we can do two quantifications by regarding the properties as sets directly.
This limited quantification can have its logic called "syllogisms". The problem is not that we don't get a universal logic, not even that we don't get a universal quantification, rather that we avoid the use of variables that is essential in math anyway.
An even more drastic avoidance of quantifications is the other old logic of "propositions".
In the Tait Logic if we don't do quantor reversals only "and" -s then it is still possible that we go back to axioms, that is or-formulas containing B , $\neg$B pairs. Allowing $\neg$ everywhere, would mean to move these down to the bottom first. But instead, we could just do one steps by:
$\neg$ (C and D) = $\neg$ C or $\neg$ D $\qquad$ $\neg$ (C or D) = $\neg$ C and $\neg$ D
Implication and other connectors can be expressed by "and" , "or" and negation, so this reversal to or-axioms is an easy way to establish all non quantific propositional truths. In fact, we might encounter or-formulas with opposing non basic states as members, so these could be regarded as axioms too. The variables and basic states are irrelevant then and we have a logic of formulas using only capital letters. An easy alternative method of establishing the derivability is the "evaluation". Write into all of the appearing letters all possible true and false combinations and calculate the truth value of the whole formula by the obvious meanings up from the bottom.
To create a derivation system that replaces these evaluations, is already useless but then to dwell upon such system as some kind of heuristic method of natural thinking is ridiculous.
But again, a side line of this propositional logic was created that introduced the meta values of "possibility" and "necessity" and lead to interesting new aspects. These "modal" logics are true directions of abstractions but can not be appreciated by first grasping the concept of variables.

Derivable Relations Of Naturals

We want to regard new basic states among the naturals, so that new axioms about these could restrict the strange models to be mere phantoms with same truths as the real naturals.
The source of these relations are ones that are visibly determined by the consecutiveness yet can not be explicitly defined by consecutiveness. The simplest example is the $<$ smaller relation.
The subjective determination is that $x < y$ means $x \lhd u \lhd v \lhd \ldots\ldots \lhd y$.
These dots of course are not allowed in a language.
The solution was suggested by Peano to avoid the dots and derive $x < y$ in a closed form.
First of all $\lhd$ implies $<$ directly. Secondly, if we already know that a number is bigger than an other then its consecutive is bigger too:

$$x \lhd y \quad \rightarrow \quad x < y$$

$$x < v \text{ and } v \lhd y \quad \rightarrow \quad x < y$$

These should not be regarded as abstract rules that are simply true, rather methods that effectively can tell all cases of $x < y$. That's why we used a new arrow and not implication.
For example, the deduction of $4 < 7$ goes like this:

$$4 \lhd 5 \quad \rightarrow \quad 4 < 5$$

$$4 < 5 \text{ and } 5 \lhd 6 \quad \rightarrow \quad 4 < 6$$

$$4 < 6 \text{ and } 6 \lhd 7 \quad \rightarrow \quad 4 < 7$$

Since in the two abstract rules the two inferred relations are the same $x < y$, we can combine them and depict the left sides or conditions as a matrix.

$$\begin{bmatrix} (x \lhd y) & \\ (x < v) & (v \lhd y) \end{bmatrix} \rightarrow \quad x < y$$

Again, the lines are alternative “or”-s and the assumptions in one line are “and”-s.
A similar method can define addition, which is just a more complex smaller bigger relation, telling how much bigger is one number than an other:

$$x \lhd z \quad \rightarrow \quad x + 1 = z$$

$$x + v = w \text{ and } v \lhd y \text{ and } w \lhd z \quad \rightarrow \quad x + y = z$$

Again, it is an effective method for all cases. For example, the derivation of $4 + 3 = 7$ is:

$$4 \lhd 5 \quad \rightarrow \quad 4 + 1 = 5$$

$$4 + 1 = 5 \text{ and } 1 \lhd 2 \text{ and } 5 \lhd 6 \quad \rightarrow \quad 4 + 2 = 6$$

$$4 + 2 = 6 \text{ and } 2 \lhd 3 \text{ and } 6 \lhd 7 \quad \rightarrow \quad 4 + 3 = 7$$

The first rule of addition can be a bit “overcomplicated” by giving it as:

$$x \lhd z \text{ and } y = 1 \quad \rightarrow \quad x + y = z$$

This way the consequences are again the same, so again we can use a matrix form:

$$\begin{bmatrix} (x \lhd z) \; (y = 1) \\ \\ (x + v = w) \; (v \lhd y) \; (w \lhd z) \end{bmatrix} \rightarrow \quad x + y = z$$

Multiplication is repeated addition, that is $x \bullet y = z$ means $x + x + \; . \; . \; . \; . \; + x = z$. But again, this is not exact, due to the dots. Peano's rules say that multiplication with 1 is keeping the number, while increasing the second member with 1 means a new addition:

$$\begin{bmatrix} (x = z) \; (y = 1) \\ \\ (x \quad v = w) \; (v \lhd y) \; (w + x = z) \end{bmatrix} \rightarrow \quad x \bullet y = z$$

Similar method gives exponentiation that is repeated multiplication.
Exponent 1 keeps the number and increasing the second member with 1 means a multiplication:

$$\begin{bmatrix} (x = z) \; (y = 1) \\ \\ (x^v = w) \; (v \lhd y) \; (w \bullet x = z) \end{bmatrix} \rightarrow \quad x^y = z$$

Observe that multiplication used addition so its matrix system contains the rules of addition too.
The complete matrix system of exponentiation contains addition and multiplication both.

We could now define a next level for repeated exponentiations, but a much smarter thing is to define all possible $x \; [n] \; y = z$ relations as quadruples.
The $n = 1$ case is addition, $n = 2$ is multiplication, $n = 3$ is exponentiation and so on.
The general rules are easy! For $n = 1$ we give addition as start, for other n we claim the usual "no change" or equality of z with x and finally the jump from a v to a next y gives a change of the w result to z by the one lower relation applied to w with x :

$$\begin{bmatrix} (n = 1) \; (y = 1) \; (x \lhd z) \\ \\ (n \neq 1) \; (y = 1) \; (x = z) \\ \\ (x \; [n] \; v = w) \; (v \lhd y) \; (m \lhd n) \; (w \; [m] \; x = z) \end{bmatrix} \rightarrow \quad x \; [n] \; y = z$$

The added beauty is that this works in itself as a single matrixed system.
Regarding these operations as relations is much cleaner. Just as we derived all $x < y$ pairs from the rules, we derive all triples for addition, multiplication and exponentiation or quadruples in this last universal case. The old fashioned view that operations produce the resulting unique value from the members as inputs is thus melting into this bigger picture of derived relations in general without inputs. The uniqueness of these last z members in these relations is merely a special feature now. In fact, the z value of addition, multiplication and exponentiation was not only unique but even more importantly it was always bigger than all used variable values. So, the possible target triples, with last z members all under a value say 10, can only be produced from values under this in the matrix lines. Thus, trying out all possible value combinations under 10 in the matrixes, we only have to derive the correct right triples and reuse these as true members in the matrixes again and again. This of course is a finite process and so we will find out concretely the left over cases. Thus all non true triples under 10 will be obtained too.
In the universal operation z is not bounding n because: $x \; [n] \; 1 = x$ for all $n \neq 1$.
But n is bounding m, so now again for every value like 10 we can try all combinations under this and will get all false quadruples too.

In general, the target relation can have arbitrary many variables and the matrix systems will derive so called "tuples" of that many numbers. Bounded systems mean that with such trial method we can list increasingly the negative of the tuple set that is all tuples missing too.
For non bounded systems, small target tuples can come from arbitrary big values of other used variables in the matrix. So by trying out all smaller values than a bound will not guarantee that we refuted all possibilities.
The derivable tuples is still a definite set with a definite complement that is left out tuples.
So this negative of the target is determined but may not be collectable mechanically at all.
We are barking at the crucial concept of derivabilities in general, that we come to very soon.
Observe also that even for bounded systems where the complement of the target is effectively determinable by the above trial method, it is still a question whether an other matrix could produce it. Usually, it is only external knowledge that helps us to find this complement matrix.
For example the negative of $x < y$ is obviously $x = y$ or $y < x$ which is easily matrixable.
The real beauty would be a mechanical negation of all matrixes. This is impossible.
Yet, amazingly, the uniqueness of z and all possible values of the other $x , y , \dots$ target variables gives an instant such method!
Indeed, if to a system with such target $R (x , y , \dots , z)$ we add the derivation rule:

$$[R (x , y , \dots , z') , z \neq z'] \rightarrow \neg R (x , y , \dots , z)$$

then we'll get exactly the complement of R as the new $\neg R$ target.
Observe that the condition of all possible other variable values was crucial! Without this we would derive correct missing values for $\neg R$ but not all of them. Indeed, the missing $x , y , \dots$ values with any z would not be obtained at all. If however R is a total function, defined for all $x , y , \dots$ independent variables then such missing tuples simply don't exist.
A much simpler added matrix derivation to a system is the dropping out of variables.
For example, adding $[R (x , y , \dots , z)] \rightarrow P (z)$ to a system defines the new target property of being any obtainable z in R.
This is exactly how we define the composite numbers as simply the results of multiplications with the added condition of neither x nor y being 1 :

$$[x \bullet y = z , x \neq 1 , y \neq 1] \rightarrow \text{Comp}(z)$$

The complement of this is the primes but we can not obtain this $\neg\text{Comp}(z)$ by adding the derivation rule $[\text{Comp}(z') , z \neq z'] \rightarrow \neg\text{Comp}(z)$. Instead we get all z values. The same would happen by trying to negate the multiplication triples and then avoid 1 values for x or y.
To derive the primes, we need to negate all compositions of z from numbers under z:

$$[1 \bullet 1 \neq z , 1 \bullet 2 \neq z , \dots , (z-1) \bullet (z-1) \neq z] \rightarrow \neg\text{Comp}(z)$$

So the dots that our matrix systems tried to avoid, sneaked back!
Luckily, we can avoid these with an other tricky introduction of new targets. Indeed, from an R old target we can obtain the $\forall u < x\ R$ relation as new T target with the added rule:

$$\left[\begin{array}{c} x = 1 \\ \\ T (u , y , \dots , z) , R (u , y , \dots , z) , u \lhd x \end{array} \right] \rightarrow T (x , y , \dots , z)$$

Using this twice for x and y in $x \bullet y \neq z$, we can obtain a $T (x , y , z)$ target that means: $\forall u < x\ \forall v < y\ (u \bullet v \neq z)$, that is, no composition of z from numbers under x and y.
And thus $T (z , z , z)$ means $\neg\text{Comp}(z)$. A natural bigger question is this:
Can we define some effective relation that we are sure can not have an effective complement?
This question is actually a dual question because we can mean it in a classical sense, that is define the relation as a number theoretical one but we can also mean it as finding a matrix system in our

new formalism and then prove that its complement is not a system. The underlying deeper question is whether all truly effective number theoretical relations are definable by our system.
The real paradox turned out to be that a classical example of non effective complement is not easy to find but such frameworks of formal systems like our matrix derivations will give an exact example of a system with no complement system inside the framework quite easily. In fact, the heuristic method of finding such system is an inner version of the deeper universality problem.
Our deeper universality question was whether all effectivities are captured by our framework. This unfortunately is not answerable for any framework. Only concrete frameworks can be compared and so external or outer universality is not provable about any framework.
The inner universality is claiming that there is a concrete system in a framework that can imitate all others. This then easily creates a "diagonal" or self referring version of this universal system that can not have a complementing system. I just sketched the whole story of effectivity but the original paradox is the real essence. No concrete classical examples but strangely easy new proofs. To some this means that this whole Effectivity direction is a baloney and doesn't relate to real math, while for others it became the real math.
To me, it is obvious that the non existence of easy examples for non effective complements is a real feature of old classical math. Not falsely created by effectivity rather brought to light. But our new math of the present is merely a window and it is indeed pretty useless for classical math.
I believe that didactical logic is the future, the avoidance of lies. And lie itself will be defined too.
The most concrete lie in New Math is the ignorance of that this fundamental paradox actually has been resolved by a grand result that came two decades after the birth of effectivity and not from the grand figures who created the field. This result is Rice's Theorem, which is distorted and presented as a trivial mere generalization of some earlier results. The lie is easy to reveal! If it is merely a trivial generalization then why did it take two decades to realize it? Simply because it is much more than a formal and fairly easily provable fact. It throws light on the basics that were not understood properly by the founders.
The bigger point is that understanding has its own logic and our present derivational logic and obsession with derivations only, is a primitive and immoral state. Because real math, both classical and new, progresses by understandings not by proofs. So, didactics is not only the moral action because we should transfer knowledge but because all creative thought is didactical and denying it or turning it into formal derivations is a deception. But we jumped ahead and will explain these later again. A historically earlier lie is how the whole matrix derivation system framework became buried under other ways to gather effective sets with effective complements.
It is ironic that the woman who made the first breakthrough, Rozsa Peter is also the one who wrote the first and last popularizing book about New Math, titled "Playing With Infinity". Yet her more known contribution, the primitive recursive functions is the solution that hides the tuple derivations as the pure start of effectivity.
These primitive recursive functions indicate three things by their name.
The functions mean totally defined functions, so the first thing that could jump into our mind is the above mentioned effectivity of the complement. Of course a function has no complement!
In spite of this, the existence of complementing pairs will be a trivial consequence for the relations defined through these functions.
The really natural advantage of functions is that they can be written into each other's variables, that is they can be composed like algebraic formulas.
The word "recursive" refers to the self usage of the target but the crucial "primitive" specifier means a destruction of the heuristic usage of the non target variables that is the lucky finds.
Only one target variable is used in such recursive manner and even that is mechanized by giving the target function at the 1 value of the chosen variable and also expressing the $n + 1$ valued function from the n valued. These initial and next value expressions of course use already defined functions from the other variables. The goal was that free choices are avoided and so the target function is determined not merely as a possible collection of all derivable cases but selfdetermined mechanically.
The formal trick that actually hides the relations, that is tuple collections or properties, that is number collections, is to regard variable values giving a fix function value. In short, an equation

$f(x, y, \ldots) = v$ with a concrete v number, collects the $x, y, \ldots$ values or tuples. Clearly, the complement tuple set is always effective too because $f(x, y, \ldots) \neq v$ is effective too. Or instead, only functions with two fix values say 1 and 2 could be regarded.
In truth, the solution was even more deceptive by regarding 0 as natural and then the 0 and 1 values give a formal digitalization of the yes and no decisions of collection. Even this choice hides a formal advantage that operating on the truth was easier this way.
So this formalization of the primitive recursive relations is the “smooth” didactical trap that hides the real power behind it. A simple fact becomes that bounded existences that is lucky finds under target variables remain primitive recursive relations. So, for example being composite is such too. The complementation means at once that bounded universalities are primitive too and so primality is a primitive recursive property. But much more complex relations are primitive too.
An interesting fact that reveals the whole situation is that for example the universal $x\ [n]\ y = z$ operation as a function of x and y is not primitive recursive as Peter herself proved it, but this same operation as a relation, as we regarded it, is a primitive recursive relation too.
So it’s not really the multiple recursion as such that alludes primitive recursion, rather the fast growing of z. If we know z then the truth or falsity of this universal operation is still primitive.
All this explains better how could the primitive recursive functions be sufficient for Gödel’s arguments about number theory, which hid even further the general effectivity lying behind.
In fact, just like no natural example existed for the non effective complement, also no easy example existed for a missing non primitive recursive relation or property! The unnatural examples for both, were actually hiding in Gödel’s results about axiom systems of number theory.
But Gödel wanted to capture general effectivity for its own sake too as an other project.
He assumed again functions as the natural form of this, just because it was formally much simpler than using relations. So instead of our matrix derivation system framework he regarded functional equation systems. This combines the advantages of both worlds. Free derivabilities and functions that can be used in each other. And indeed, all possible effective functions are obtained.
But this success is misleading. We derive the individual cases of $f(a, b, \ldots) = v$ but these define an effective f only if uniqueness and totality stands for all the derivable cases of f.
None of the two can be of course seen from the individual cases.
Kleene realized that uniqueness can be guaranteed quite easily if avoid the free derivation idea and follow the deterministic line of the primitive recursive functions. Amazingly, only a single new rule has to be added. This keeps uniqueness but opens the gate of non totality.
While primitive recursion used a single variable’s recursion to avoid lucky finds, here one variable’s lucky find is portrayed as checking its values for all naturals increasingly until one is found where the function takes up a value. Again, this value is chosen as 0. This then is called an “unbounded search” that determines the new function value for the other variable values.
Of course, such search may not be successful ever and at such other variable values the function is not defined. He called these non total functions “partial”. “Accidentally” of course we’ll have partial functions that are in fact total because the searches are successful for all independent variable value combinations. But these total ones are not formally recognizable. Yet, he showed that all total functions obtainable from Gödel’s systems are obtained here too.
To understand all this is quite easy:
The effectively collectable tuple sets are not all having effective complements! In fact, the natural is to have a non effective complement. That’s why non total functions must be accepted.
But those tuple sets that do have complements inside are not only “rare” in a framework but actually not even collectable effectively. That’s why the total functions are not recognizable from their forms in Kleene’s system.
The real important didactical fact is that we only had to go through functions in order to use the primitive recursive relations that contain almost all effective ones with effective complement.
This subjective “almost”, becomes objective in the details of Kleene’s proof by producing all partial recursive functions $f(n)$ and similarly all of Gödel’s too, from fix $t(e, n, x)$ and $u(x)$ primitive recursive ones. Namely, a single unbounded search of $t(e, n, x) = 0$ in its x variable gives the $\mu\ t(e, n)$ two variable partial function and then applying u to its value gives $f(n)$. The remaining e parameter is thus a code number of f in this uniform method.

More generally, e could be regarded as a code for a system, while x as the code number of a derivation of $f(n)$ by the system. The minimal choice is then mostly just an existing choice for x and u simply detaches the derived final $f(n)$ value from the derivation coded by x.
My nomenclature is to call the derivable, that is effectively collectable tuple sets that do have similar complement, as totally derivable or totally effective. This nicely coincides with the functional name as total being one that is defined for all tuples.
The mentioned paradox that we don't have easy examples of non effective complements might be the reason for the mentioned early totality obsession and arising false view of totality included in Effectivity. Then, the non total effectivity was called "partial" for functions and for sets or tuple sets the idiotic name of recursively enumerable was introduced. Everybody knows today that these are the real fundamental effective sets but nobody cared to rename the concepts accordingly.
All this is even more evident from Turing's heuristic machine approach that can be easily approached from the tuple sets too. Indeed, the naturals can be viewed as mere repetition of 1-s while the commas separating them as 0-s. So a tuple is simply a binary string with fix many 0-s.
Using more than two symbols unrestrictedly means an alphabetical text that replaces tuples. This includes mixed, that is different long tuples at once. Effectivity defined for texts is thus a bigger picture that uses varying long tuples in disguise. The more intricate detail is that Turing's system was not relational, reducing the tuples to the consecutiveness relation, rather truly alphabetical as local alterations. That corresponds to how elementary school kids calculate digit by digit.
Now we return to our clearly naïve matrix method that failed to formalize effectivity.
We can easily state any of these matrix derivation systems as new axioms that introduce the target relation. The hidden quantification is quite simple. Indeed, the target variables are universal and the rest of the helping variables in the matrix are merely assumed as lucky finds, that is they are existential. So a matrix derivation system:

$$\begin{bmatrix} x, y, \dots, z \\ \\ u, v, \dots, w \end{bmatrix} \rightarrow R(x, y, \dots, z)$$

means: $\forall x \forall y \dots \forall z \, (\exists u \exists v \dots \exists w \, [\quad] \rightarrow R(x, y, \dots, z))$

The machinery of Logic can then derive the cases perfectly too.

The Induction Axioms

Our matrix derivation method started from $x \lhd u \lhd v \lhd \ldots\ldots \lhd y$ as $x < y$ and even the operations are naively definable as "dotted" repetitions of same members.
These are of course not allowed formulas and thus not proper definitions either.
The matrixes determine the cases exactly and so actually define the relations but this is not the accepted definition of definition itself. Only truly explicit formulas are usable. For example, $<$ is truly definable from the basic relation of addition as: $x < y \ := \ \exists u \ (x + u = y)$.
Obviously, $\forall x \ \forall y \ (\exists u \ (x + u = y) \rightarrow x < y)$ is true too and so we can derive the cases of $<$ from $+$ by the $[x + u = y] \rightarrow x < y$ matrix rule too.
But the crucial extra feature here is that the $<$ target is not appearing in the matrix, so we never use earlier cases to derive new ones. In normal matrix derivations that was the whole point.
Such self using, case by case determination with derivation systems is not regarded as explicit definition but can be described by the word effective too.
So Explicit and Effective are totally different directions in math though in everyday language they are similar. "I told you explicitly" or "I told you effectively" mean the same.
In math, Explicit is simply being logically expressed that is being formula while Effective is the crucial new mechanical feature that emerged after Gödel's results became examined.
But now we have to talk about a duality of determination first:
Just because these derivable relations are so well determined by the matrixes, it doesn't mean that their properties are determined by these matrixes too.
Addition and multiplication for example have the well known exchangeable order property.
$2 + 3 = 5$ and $3 + 2 = 5$ is a general pattern, in other words: $x + y = z \rightarrow y + x = z$.
And similar is true for products. But our rules merely derive by increments in the second member. So these two symmetries are not obvious at all.
And indeed, the exponentiation defies such symmetry: $2^3 = 8$ but $3^2 = 9$.
If every single such derivable relation has its own secret world of truths then how can we penetrate all of them? Because we have a "magic wand". Just as the tuple sets are derived cases by case, the "cases" of such universal relations could also be used to derive the universality. Namely:
If there are some initially true cases and the cases inherit each other and we can prove this inheritance, then we can claim a universality of the feature.
This of course means to claim this jump to universality from inheritance for all possible relations.
A problem is that a relation has many variables and so a concept of inheritance is not clear at all.
On contraire, for a simple $P(x)$ property the idea is perfect.
If $P(x)$ is true for $x = 1$ and is inheriting from every x to $x + 1$.
That is: $P(1)$ and $P(x) \rightarrow P(x+1)$.

Then our axiom should be: $[\ P(1) \text{ and } \forall x (P(x) \rightarrow P(x+1)) \ ] \rightarrow \ \forall x \ P(x)$.

The heuristic idea is simply to use this primitive property idea for relations too, by merely regarding the variables one by one. This of course should make the proofs of universalities for relational properties quite difficult because we have to work through all variables.
And indeed, already to prove the $x + y = y + x$ feature is quite tricky even with using this simplified functional notation.
Our addition derivation rule merely claims now that: $x + (y + 1) = (x + y) + 1$.
First we use inheritance to show that the $x + y = y + x$ exchangeability is true for $x = 1$.
That is : $1 + y = y + 1$. The $y = 1$ case is trivial: $1 + 1 = 1 + 1$.
The inheritance from y to $(y + 1)$ can be obtained as :

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

The first equality used the addition derivation with $x = 1$, the second used the y assumption.
Next we show that a first member derivation is true for addition: $(x + 1) + y = (x + y) + 1$.
For this, the $y = 1$ case is trivial again : $(x + 1) + 1 = (x + 1) + 1$ and then the jump is:

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

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

The first equality used addition derivation, the second used the y assumption, and the third again addition derivation in reverse. So finally the exchangeability itself is $x+y = y+x$ which we already proved for $y = 1$ and the inheritance means:

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

The first equality used addition derivation, the second the y assumption and the third the reverse of addition derivation from first member.

As we see, this is a nightmare. And indeed, we think in inheritances of relations not variables.
If we have an $R(x, y, \ldots)$ relation that we suspect to be universally true then of course we can not go in steps of 1 with all variables simultaneously to cover all tuples.
We can still go in steps of 1 by inheriting R from all variables being under a number say 10, to the variables being under or equal to 10. So the inheritance for this 10 value would be:

$[(x, y, \ldots < 10 \text{ and } R(x, y, \ldots)] \rightarrow [(x, y, \ldots \leq 10 \text{ and } R(x, y, \ldots)]$

We can use a new variable instead of 10 and state the same in general but it is very rare that this restricted inheritance is provable that is successful. A more promising possibility would be to replace the new variable as stepping number, with an f function calculated from $x, y, \ldots$
But it is still very restrictive that all variables must stay under same common steps.
The real freedom is to allow a sequence of $R_1, R_2, \ldots$ stepping relations.
These should be widening and eventually cover all possible tuples.
The induction axioms can be thus quite complex but their validity for the real naturals is clear.
Most importantly, the garbage extra numbers then must obey these universal features too.
It can’t stop their existence but it surely puts a strain on their weirdness. It is hard even to visualize how these double infinite snakes could behave as naturals with more and more such universal features required to be true about them.
For example the simplest derivable relation $<$ will have the inductively provable universality:
$x \neq y \rightarrow x<y \text{ or } y<x.$
Claiming this as axiom, at once means that the garbage numbers can not just float unrelated to our names, rather have to line up in $<$ relation to those and each other. The weird models become ordered sets continuing the real naturals. Then the addition, multiplication and so on, create even more intricate internal structure in the garbage. Each will have infinite many features to be obeyed simply because those features inherit among the names. In a sense we are our own enemy because the more we require, the more intricate the garbage becomes.
But remember the idea was that requiring enough, we could turn the garbage into a ghost world that will have only the same truths as the names.

A Classical Example, The Infinity Of Primes

We already used the infinity of primes as example for our matrix form. We replaced the actual claim of infinity with the claim that there are arbitrary big primes. That is for any number there is bigger prime. This is already a cheating because the actual infinity would involve that we pick one by one bigger and bigger primes as a set. Of course we want to avoid talking about sets.
But now I will go into the actual proof of Euclid and amazingly we'll see that even deeper problems arise with translating the proof into exact logic.
The crucial idea behind the proof is the combining of two lines of reasoning.
One is producing a sure prime for any n number but not above n rather up to n.
In a sense, we even strive for smallness, looking for the smallest non 1 divider of n.
Every n number has trivially two dividers, namely 1 and n itself. So, excluding 1 still leaves at least one non 1 divider or so called factor. Among these factors, there has to be a smallest that we can even abbreviate as: minfact (n).
Composites are n numbers that can be composed, that is made as a product from two smaller numbers. So, only factors can be used because the trivial 1 divider must be multiplied with n.
Using a smaller factor automatically guarantees an other to multiply with. Thus minfact (n) tells the compositeness or primeness of n beautifully. Namely, if minfact (n) is smaller than n then and only then is n a composite, while if n is a prime then minfact (n) $=$ n the prime itself.
Our first claim is that minfact (n) is always a prime even if n isn't.
To see this is amazingly simple indirectly. Indeed, if minfact (n) were a composite, that is having some f smaller factor, then f would divide n too and so minfact (n) couldn't be the minimal factor because f were a smaller.
Now that we know minfact (n) is always a prime, we'll use this but not for n rather for a much bigger N number created from n, so that minfact (N) is indeed bigger than n.
One sure way of having an N that minfact (N) is bigger than n is if we know that none of the non 1 numbers up to n that is none of $2 , 3 , \ldots , n$ divide N.
To create such N is again very heuristic! Namely, we create a number that definitely has these as factors and then we simply add 1 to it. This makes a trivial 1 remainder and thus a non dividability. So, we only need a number that is dividable by all of $2 , 3 , \ldots , n$.
Well that's simple, we just have to multiply these together. The really stupid name for this product is "factorial" and abbreviated with exclamation mark. So $n\,! = 2 \bullet 3 \bullet \ldots \bullet n$
Thus $N = n\,! + 1$ and so minfact (n ! + 1) is a prime above n.

The most obvious problem with the proof is that the factorial is not an explicit definition. It used dots again. To use Peano's idea, that is a matrix derivation system is easy:

$$\left[\begin{array}{c} x = 1 \;,\; z = 1 \\ \\ u\,! = w \;,\; u \lhd x \;,\; w \bullet x = z \end{array}\right] \rightarrow \quad x\,! = z$$

This of course means to introduce the factorial as a new basic relation and convert this matrix system into a new axiom. The induction axioms then claim an infinity of new facts too.
But the claim of the infinity of primes relies only on multiplication. Is this a delusion? That is, the multiplication is not enough and hides the bigger needs of the derivation? Yes, in part this is true but not in the sense we suggested it. Namely, the induction is the only extra thing we need beyond the multiplication or about the multiplication. Because the factorial can be eliminated completely.
Indeed, instead of naming exactly this new prime above n as minfact (n ! + 1) , we can merely prove that there is a number that is dividable by all numbers up to n. This goes smoothly with induction using multiplication! Then from this fact we can derive the existence of such prime too.
So as we see, induction appears to be "all powerful" but requires tricky translations of the common sense classical ideas. Only a higher view shows that induction is not all powerful either.

Gödel's Incompleteness Theorem And Effectivity

If the names, that is the real naturals are a model of an axiom system of number theory and all other models have the same truths in them, then for every statement or its opposite exactly one of them must be a derivable theorem, namely the one that is true in the models.
Indeed, one of the opposing statements is true among the naturals. If the opposite were theorem then the naturals weren't a model at all. If none of them were theorems then by the completeness of Logic we had models where either is true. The reverse is even more obvious by simply models obeying Logic. Indeed, if every statement is decidable that is itself or its opposite is theorem then all models share these as truths. So, to avoid "alternative models" that is models with alternative truths, is quite simple theoretically by selecting all statements or its negatives as axioms.
Of course, we could easily make contradictory choices, that is making the total set inconsistent.
A seemingly better but just as useless idea is to choose as axioms from all pairs of opposite statements the ones that are true among the real naturals. Then we wouldn't even need Logic.
Unfortunately, we can not recognize the truth of the real naturals for all statements.
Indeed, universal statements can only be verified by checking infinite many numbers.
The induction axioms are definitely recognizable because they are generated by simple rules.
Unfortunately, the induction axioms don't work either!
By inheritable conditions we can not claim enough universalities that decide every statement.
Amazingly, already in this section, I will make this claim very plausible.
It was the same Gödel who proved the Completeness of Logic who also proved this unavoidability of undecidable statements, that is alternative models. This was the already mentioned new wisdom that Gödel revealed in the intermission of the 1930 congress to the few understanding ears. As I mentioned, this was not approached from models. So, the concept of alternative models, was not used at all, and not only in Gödel's head but even in the following decades. Such model approach should exist and finally became examined, but why it is not a natural approach is still a mystery.
Quite oppositely, a crucial factor did become clearer and clearer already in the following years after Gödel's discovery. This factor is Effectivity. Above I was very naïve and rejected the truth idea to collect the axioms simply because "some" truths are only "recognizable" in infinite time.
But there is a deeper level! Namely, even if some oracle would give the truths of finite many individual statements eventually, we wouldn't have them all being recognized by a single system.
On contraire, the induction axioms are all generated by same rules. The theorems then are derived by Logical rules again. So, only when the axioms are given by some rules, are really the set of the theorems a derivable set. This new vision that we regard derivations of any objects by any rules, that is derivable sets in general, is the crucial new point.
But we also used the words: "recognizing" and "generating".
Recognizing seems opposite to derivation because we start with the object not get it as result.
Generation is deterministic while the derivations involve free choices.
And yet, all these differences are immaterial!
Behind derivation, recognition, generation, lie a common reality that I called already Effectivity.
Nowadays a trend is to call it "computability". The greatest champion of the post Gödel clarification years, Alan Turing regarded Effectivity as actual machine recognition or generation and defined an abstract computer model. In spite of this, the term computability is false.
The duality of the machine usage as recognition of an input or generation of an output shows that the above confusions are now built in as alternative applications. It is also a strange coincidence that already earlier we saw this duality in Logic itself. Indeed the normal vision of deriving the theorems as outputs is identical with deriving contradictions from their negatives as inputs. But this analogy is not quite correct because these are both derivation systems and thus contain free choices unlike machines. An instant idea to avoid the freedom of choice is to try all possible choices in some fix order. This feels like a vital intrusion on "derivability" but we'll see that for our purpose of showing the necessity of undecidable statements it doesn't matter.
Free choice using derivations, that at first don't even fit into a computer or machine, are more akin to games. Chess for example has fix rules. The players have to obey the rules but have free

choices. This analogy will be used to show the promised plausibility and it will also show why the free choices can be irrelevant even here.

We put some pieces on the chess board and then call in a champion and ask his opinion.

Probably he finds the situation on the board alarming or as a result of two lunatics playing against each other. But then we state our actual question. Is this situation possible if the two lunatics played by the rules? A champion will be able to recognize situations not only as possible but maybe even recreate the history of the game that brought it about.

To tell impossible situations is much harder!

I am not a good chess player but I know that the bishops remain on the same color. So, I know for sure that if the two bishops of a player are on same colored square then this situation is phony, impossible by the rules. The other pieces interrelate more intricately. But most crucial is this:

While the possible situations are all verifiable by a discovered history that leads to them, an impossibility is only verifiable by examining all possible games! So the set of all possible situations and its complement, the set of all impossible situations are not symmetrical. This second is much more complex.

The derivations of theorems from finite many axioms by Logic is quite similar as deriving situations on the chess board. If we have infinite many axioms derived by some rules themselves, then these rules plus Logic will give the total rules that derive the theorems. But again, the derivable theorems is a set collected by finitely verifiable derivations while the complement set, “the left over” is a much more complex set. This has elements that can only be refuted by examining infinite many possible derivations. We claim that this individual complexity of the elements in the complement is leading to a total complexity of the whole complement as a set. Namely, unlike the derivable set that is collected by a single system of rules, the complement is not collectable perfectly by any rules. The word “perfectly” here simply emphasizes that we don’t merely want to collect all elements but exactly only the elements. The word “collection” of course should mean this anyway. This uncollectability of the complement of a collectable set is our plausibility. We don’t claim this as a universal necessity but merely say that if the collection rules are complex enough, then the even more complex complement set could be an uncollectible, that is unruled set. So for example at the chess, the impossible situations might be still describable that is collectable by some rules because the game rules themselves are fairly simple. But we claim that using a complex enough game, the impossible situations are not ruled at all.

Now comes a vital difference between games and Logic:

Unlike the chess situations that do not form any obvious relationships except their evolutions in the histories that create them, the statements have formal relations, most importantly they form pairs of being negatives of each other. This is the cause that creates the false expectation that we could derive exactly one of each pair with some good or rich enough axiom system!

But if we could derive exactly one of each pair then the complement set would be simply the set of all other members of the pairs. So this would be derivable instantly too, contradicting that it has to be a much more complex set that is not derivable by any system at all. The only way the complement of the derivable statements or theorems, that is the non theorems as a set can be a complex non derivable set in general, if the theorems leave some pairs undecidable, that is not being either member in them.

As we see, this plausibility was a bit vague and relied on having a complex enough collection system already in the first place. The formal arguments all emphasize this side. They prove from the collection system that the complement can not be collectable. These arguments are indirect and are also called as diagonal or self reference arguments. This overshadowed the main and plausible Effectivity argument that I tried to explain above. Also it explains why Gödel himself missed the Effectivity under his own proof and emphasized the role of language.

A final explanation is owed for the title itself. This necessity of undecidable statements in axiom systems is called the Incompleteness Theorem because we simply call an axiom system complete if all statements are decided by it. Clearly, this Incompleteness Theorem is not an opposite of the Completeness Theorem. That was talking about the completeness of Logic while this is talking about the incompleteness of most or rich enough axiom systems.

Gödel's Fundamental Realization, The Power Of Multiplication

The derivation rules give only the actual cases for $<$, addition, multiplication and so on.
The universal features are not derivable. More precisely, only universal inheritances are derivable. Luckily, from these we can conclude universal relations using formally the infinite many induction axioms stated for all possible relations. The real trouble is to find inheritance steps.
This is not a formal trouble because restricted steps can prove the more tricky ones.
It is not quite true that only inheritances are the provable universalities from the cases.
It is true that only relative, implicative universalities are provable from the case deriving axioms.
But the different derivable relations can imply each other too.
The simplest, first derivable relation $<$ can be used as a special condition.
This is called bounded universality like: $\forall\ x\ [\ x < 10 \rightarrow\ \ .\ .\ .\ ]$.
We can even use a variable instead of 10 and still, for every value we have an upper bound.

If all universalities are bounded then the statement can be proved from the derivational axioms of the appearing relations and a few universalities of $<$. No induction is necessary!

Now, this fact suggests again that more and more derivable relations could be useful.
But we are wrong! Gödel proved his incompleteness for the number theory with only addition, multiplication and their infinite many induction axioms. This of course could mean that that's all he could achieve and so using higher derivable relations like exponentiation are nor necessarily incomplete. But this is not the case and that follows from the details of his proof. The derivable relations are irrelevant after multiplication. They can not create completeness. But how can we know this about all of them? They are infinite many and seemingly infinitely different. Their truths are not mechanical at all, as the non exchangeability of the exponentiation order shows. Only the infinite many induction axioms can grasp these differences.
But an even more shocking second level is this: The induction axioms are irrelevant too.
So, our above mentioned induction avoidability for bounded universalities is sufficient to avoid induction of even addition and multiplication from Gödel's original proof.
This sounds straight out crazy, because without the induction axioms we obviously know that the system is incomplete and for example even the exchangeabilities of $+$ and $\bullet$ are underivable.
So we get something trivial by getting rid of the induction and achieve nothing.
And indeed this crazy idea makes only sense from a third level:
This third level shows that Gödel's heuristic idea that showed why the higher derivative relations like exponentiation are irrelevant, works for any axiom extensions not only derivative rules as new axioms. So adding any new axioms to the simple system of $+$ and $\bullet$ without induction, can never be complete. And this then includes the inductions as merely special axioms.
This sounds fantastic but still begs the question what was the new concept actually derived about this obviously incomplete system of $+$ and $\bullet$ that could then be inherited to all extensions and also implied incompleteness? Indeed, this is the essence and it is simply that the non theorems are a non derivable set of objects.
But Gödel's crucial result seemingly talked about something else, namely it claimed that:

All derivable relations are explicitly expressible with addition and multiplication!.
In particular, exponentiation is explicit too! That is:
Exponentiation is definable from addition and multiplication with a single formula.

In elementary school we teach multiplication as repeated addition and exponentiation as repeated multiplication. I think in high school it's time to tell that this repeatedness is actually involving dots and going "outside" to tell the number of members, so it is not really explicit.
Finally, in tertiary level it should be explained how multiplication is a universal de-coder and so can define exponentiation explicitly too.
The basic idea is very simple. We can order a unique number or c code to any tuples of numbers. This can not be explicit for all tuples because the lengths of the tuples are themselves different, so

in a sense they themselves contain dots. For one fix length of tuples of course we can have explicit formula but we need better than that. Amazingly, the reverse de-coder can be explicit for all tuples because we only have to obtain the particular tuple elements.
So a $T(c, i)$ function could give the i-th element of the c coded tuple.

Then for example, the $x^y = z$ exponentiation simply becomes the existence of a tuple with the repeated multiplications with x, that is the x powers. So, the claim using c is:

$$\exists c \, \forall i < y \, [\; T(c, 1) = x \;,\; T(c, i) \bullet x = T(c, i+1) \;,\; T(c, y) = z \;].$$

Observe two things! Firstly, if instead of a single code, we used more, say two M, N numbers, that is a $T(M, N, i)$ function, it would work just as well.
Our explicit form would merely start as $\exists M \, \exists N \, \forall i < y \; . \; . \; .$
Secondly, our claiming of these existing tuples are not unique even if T is. Indeed, for example the longer tuples that contain the same x powers but after y many such, contain other arbitrary elements, also perfectly define our exponentiation. And this means that we don't really need a T function that gives all tuples, merely a function that contains all tuples as the beginning values.
In fact, the possible i values for which T is defined at a particular M, N values, doesn't even have to be all consecutive numbers up to an n number. Of course T must contain longer and longer such full beginnings in its i domain.
This visual description was vital to see behind the following formal claim that sums up our goal:

There exists a $T(M, N, i) = t$ formula, so that for any $(t_1, t_2, \ldots, t_n)$ tuple, there are M, N numbers with which this tuple is a beginning of the function determined by T.
So, for any $(t_1, t_2, \ldots, t_n)$ we have: $\exists M \, \exists N \, \forall i \, [\; i \le n \;\rightarrow\; T(M, N, i) = t_i \;]$.

There are two ways to show this. Both require some beautiful applications of basic number theory:
A product of two numbers is trivial if one of them is 1. So, $1 \bullet n = n$ or $n \bullet 1 = n$ are trivial.
In the same sense, the 1 and n itself are the trivial dividers of an n number.
A c number is composite if it is a non trivial product: $a \bullet b = c$
The non composite numbers are $1, 2, 3, 5, 7, 11, 13, 17, \ldots$
The first 1 is "trivially" non composite because its only divider is 1.
The other non trivial non composites are also called the prime numbers.
To exclude 1 from the primes has its most obvious reason as follows:
First, we can realize that all numbers can be written as products of primes.
Indeed, we simply have to write all composites as non trivial products and then the factors again and again. These factors are decreasing, they are never 1, so they have to become primes.
The surprising fact is that, no matter in what order we go, the final prime factors are the same.
If we allowed 1 as prime, then clearly such uniqueness were impossible, because extra 1 factors could be added. But there is a different line of logic for excluding 1 from the primes too:
Two numbers are relative primes if they have no common divider, except the trivial 1.
Here we don't exclude 1 as potential relative prime. In fact, it is relative prime to all numbers.
And yet, the exclusion of 1 from the primes is supported by the relative primes too, because:
Two numbers are relative prime, if and only if they have no common prime factor.
This obviously follows from being relative prime, but the reverse only because 1 is not a prime.

Now we can claim the basic theorem that leads to our first approach to $T(M, N, i)$:
For any $\{ a < b < c < \ldots < m \}$ finite set of numbers, there are M, N numbers that the $S(M, N) = \{ z \;;\; zM+1 \text{ divides } N \}$ set is $\{ a < b < c < \ldots < m < \ldots \}$.
So S contains our finite set as "beginning" that is as smallest elements.
By the way, this S is a finite set too because there are only finite many dividers of an N.
For the proof let $M = m! = 2 \bullet 3 \bullet \ldots \bullet m$.
Then for every z that divides M, the $zM + 1$ values are all relative primes.
First of all they can not have a non 1, M divider as divider, because such leaves 1 remainder.

Now if two of them $j M + 1$ and $k M + 1$ had a common p prime factor then p would divide their difference $(k - j) M$ too. p divides separately and can't divide M since it is an M divider. But neither can divide $k - j < m$ because $M = m!$, so $k - j$ is an M divider too.
The suitable N is $N = (a M + 1)(b M + 1) \dots (m M + 1)$.
Obviously then the $a, b, c, \dots, m$ numbers are all in the S set. All we need to see is that up to m, no others can be in S. But for all other z values, $z M + 1$ is relative prime to all the $a M + 1, b M + 1, \dots, m M + 1$ numbers and so their product cant be dividable by $z M + 1$.

Let a pair coding $< a , b >$, be unique and also monotone in its first member, that is:
$< a , b > \; < \; < a' , b > \; \rightarrow \; a < a'$ and lets define a $T (M , N) = (f_1 , f_2 , \dots)$ function by our previous $S (M , N)$ and this $< >$ by: f_i is the first x number that $< x , i >$ is in S.
Our claim is now what we already foretold:

For our every $(t_1 , t_2 , \dots , t_n)$ tuple, there are M , N numbers that T starts as our tuple.

Indeed, first create the pair code set of our tuple, that is $\{ < t_1, 1 > , < t_2 , 2 > , \dots , < t_n , n > \}$.
Or in increasing order as $\{ a < b < c < \dots < m \}$.
Then find M , N so that $S (M , N)$ contains this code set as beginning or smallest elements.
The numbers in S after m, that is being bigger than all our codes, can by monotony only be codes of i indexes under n with first members of the pairs being bigger too than ours.
So, up to n our tuple members will be the first choices.
To turn this into a formula, we replace S with its relation " $z M + 1$ divides N " but of course with $z = < x , i >$. So:

$T (M , N , i) = t \;\; \leftrightarrow \;\; t =$ first x that $< x , i > M + 1$ divides N Or to be even more exact:

$$< t , i > M + 1 \text{ divides } N \text{ and } \forall x < t \; (< x , i > M + 1 \text{ doesn't divide } N)$$

The minimalization quantifier bounded by t, guarantees that this formula is a totally derivable relation. It is derivable and its complement is derivable too. Also, its derivation doesn't require induction axioms. The derivational axioms of addition and multiplication are enough, with the infinite many name axioms of consecutiveness of course.

An easy pair coding $< x , i >$ that can be used, is : $(x + i)^2 + x$.
For its uniqueness assume that $< x , i > = < y , j >$ and also suppose that $x + i < y + j$,
that is $x + i + 1 \leq y + j$ were true too. Then:
$< x , i > = (x + i)^2 + x < (x + i)^2 + 2 (x + i) + 1 = (x + i + 1)^2 \leq (y + j)^2 < < y , j >$
Contradicting our assumption. Similarly, $x + i > y + j$ would contradict it too, so $x + i = y + j$.
But this with $< x , i > = < y , j >$ at once implies $x = y$ and this also $i = j$.
It's also obvious that this pair coding is monotone in both members.

Observe that, in spite of its simplicity, this T is exactly as we described it at the start.
It's not defined for all numbers up to a point. It will be defined for up to longer and longer points as beginning of its domain but after these, quite erratically. Depending on what i values there have a t so that $< t , i >$ is in S. For some M , N values there will be no possible i at all.
Since $T (M , N , i) = t$ is a totally effective relation, it would be easy to extend T into a full domained function that is for all M , N , i natural numbers. Formalist textbooks proceed with these, right from the start, intentionally hiding the "beautiful ugliness" of the real T.

The second approach avoids $S (M , N)$ and $< >$, rather gets a T directly.
In a sense it is quite opposite to our first approach and yet very similar. It doesn't collect those z numbers for which $z M + 1$ divides N, rather the remainders themselves of all $i M + 1$ in N.

Above, in the S set collection, $z M + 1$ being divider clearly would always allow a $z = 0$ value, so it was natural to regard only real naturals, not including 0. But then the paircoding could easily allow $< 0 , i >$ non zero codes, so 0 could have been included in tuples too. We could have even asked about negative numbers allowed. These are healthy abstractions.
But most books regard even the i index from 0, which is abstraction abused. It is Formalism!
Nobody thinks 0 as the starting member. Only maybe a genius creating new forms!
Peano, who formulated the derivation rules of addition, multiplication and exponentiation observed already that starting from 0 does formally simplify the rules. In fact, the anecdote is that once at the opera when they gathered their coats after the show, his wife said "did you get all pieces?". Peano said "yes, zero, one, two, three". The wife looked at the four coats and said "and you are the mathematician". Yet, just as Hilbert, Peano wasn't a Formalist either.
But returning to our second approach, the zero is a very natural concept as remainder when in fact we have dividability, that is no real remainder. And the oppositeness will not mean that we have to exclude these. In fact they fit in perfectly with all other remainders, so the $(t_1 , t_2 , \ldots , t_n)$ tuples will be generated as remainders and they can include 0-s.
The theorem we use is called the Chinese Remainder Theorem:

Let $s_1 , s_2 , \ldots , s_n$ be real non zero naturals and all relative primes to each other!
Then for every $t_1 , t_2 , \ldots , t_n$ values, each under the corresponding s, that is $t_i < s_i$,
we can find a number N so that these t-s are all the remainders of the corresponding s-s in N.
And these t values may include zeroes just as the remainders do.

As we revealed, we use the same $i M + 1$ construction but now for all $i = 1 , 2 , \ldots , n$ and these will be used as $s_1 , s_2 , \ldots , s_n$ of the Chinese Remainder Theorem.
So this Chinese Remainder Theorem defines our N now, but our M is a bit different too.
It's again an $m!$ but this m has to be bigger than not only all members but the n number of elements too. So a simple choice is $m = t_1 + t_2 + \ldots + t_n + n$.
The first condition of the theorem about the values being under is then trivially true because :
$t_i < m < m! = M < i M + 1 = s_i$
The other condition, the relative primness is again easy to see the same way, except now k and j are not any two z values that divide M, rather indexes up to n. Correspondingly, now $(k - j)$ is an M divider not because it is under the m maximal number, rather because it is under $n < m$.
The final explicit formula for this direct $T (M , N , i) = t$ is :

" t is the remainder of $i M + 1$ in N " : $t < i M + 1$ and $\exists\, q < N \, [N = q (i M + 1) + t]$

Finally, the proof of the Chinese Remainder Theorem is itself a total opposite in how we created the N number before. Indeed, it was simply the product $(a M + 1) (b M + 1) \ldots (m M + 1)$.
But now it is not given "explicitly", merely its existence is proved, though ingeniously, still involving the product $P = s_1 \, s_2 \ldots s_n$.
The $0 , 1 , 2 , \ldots , P - 1$ values under P mean P many choices to try as N and as easy to see, they all give different remainder tuples.
But the possible under valued tuple combinations are also this many combinatorically as product of choice numbers. Thus, at least one tried value under P must be an N, giving our tuple.

This direct $T (M , N , i)$ is always defined for all i values up to a point, namely when $i M + 1$ exceeds N. Some people define remainders even beyond there, by regarding the q quotient as 0.
I don't think that this is as natural as the remainder to be 0.
But as we can see, a full extension is much more natural for this T .

The Cause Of Incompleteness Again

After Gödel's discovery, the minimal language of arithmetic is clear: $\triangleleft$, $<$, $+$, $\bullet$
Exponentiation and beyond is not necessary, we only need these five relations as basic.
This is not quite true. First of all we need $=$, this mostly logical relation and on the other hand we don't really need $\triangleleft$ and $<$ because they can be defined explicitly from $+$.
If we want to use special features of the bounded universalities then $<$ can be still kept as basic.
This whole linguistic view is totally opposite to the section we had before the previous, where we outlined a heuristic vision of derivabilities already as a general cause of Incompleteness.
Gödel regarded incompleteness as a language barrier. As we'll see in some sense it is.
But the other side, derivability is still the main issue, the cause of incompleteness.
The simplest way to tell the concrete cause of incompleteness is by these two facts:

1.)
If in a universe of objects we have a perfect framework of systems that can derive objects and one system is rich enough, so that other systems can be modeled inside, then the complement of the derivable objects by this system, that is the underivable or left out objects are not only left out or underived by this particular system but it is underivable by any system at all.
There is no system that could derive them and only them exactly.

2.)
If in a universe of statements, that is in a language, an axiom system can express numbers and grasp the arithmetic of $+$, $\bullet$, then the derivation of such number cases of relations or already properties, can imitate all derivation systems. This of course could be caused by extremely complicated set of axioms, like accepting all truths about numbers with $+$ and $\bullet$.
But if this is not the case, namely the axioms themselves are derivable by some rules too, then the set of all theorems is also a derivable set. But by 1.), the set of non theorems is not derivable.
This instantly implies that the system can not be complete, namely there are undecidable statements. Indeed, grasping arithmetic also means that our system is consistent and so there are statements that are not theorems. Namely, the negatives of the theorems are all such intentional non theorems. If our system were complete, that is for every B statement exactly one of B or $\neg B$ were theorem then these intentional non theorems were all the non theorems. But then a simple derivation of exactly the non theorems were possible by using the system itself to derive the theorems and then apply a formal negation, that is create the intentional non theorems.

Some would argue against this two stage explanation and claim that only the second part is the point. In fact, they would point out that the axiom system being able to imitate all systems as case theorems means also that it can imitate its own derivations inside as special derivations and so the language can talk about its own derivations. In a sense it can form a statement which actually says that itself is not derivable. So the real cause is this self reference.
At the present it is not absolutely sure whether this view is wrong or not.
We'll see pro and con features. But the crucial point is this:
In the universe of statements the formal negation creates a pairing of the objects.
Derivability doesn't rely on such formal splitting.
If it could be proven that all the object derivabilities actually hide such duality, then the self reference could be stated as the fundamental cause. But this doesn't seem to be true.
So, axioms and Logical derivabilities seem to be merely special derivable sets where simply a formal duality of truth and falsity is also assumed. This assumption is far from simple though.
In fact, very tricky because they don't tell which are the truths and which are the falsities, instead they create only an operation of negation. Still they try to be consistent by not to derive any two or in more general any finite many contradictory statements.
Amazingly, just as this consistency was enough to have models by Gödel's Completeness Theorem, this consistency is also enough to be incomplete by improving his Incompleteness Theorem.

The Cause Of Underivability, Part 1, Not Revealing Yet

A formal “logic” behind the fact that imitation implies underivable complement could be this:
If the system can imitate all others, then the complement should be a “complement” of all possible systems and thus being none of them.
Of course, one set can’t be a complement of different sets, so it is indeed a purely formal and stupid argument. More importantly, we see no plausibility why the complement of a set derived by a rich imitating system couldn’t be derived by an other rich system.
At present, we don’t have a truly plausible explanation. The reason this wasn’t missed, is because the actual proof of the underivable complement turned out to be such a simple and reoccurring heuristic method. Almost a primordial self reference. So, even those who don’t accept the axiomatic self reference as basic, rather the underivable complements, actually in the proofs of this, will accept a simpler form of self reference.
We’ll come to this simpler self reference called “anti diagonality” later.
But there is an interesting dilemma arising from the rich systems that did become explained and which I regard as absolutely fundamental. That’s what I claim as the cause of underivability.
The dilemma is this: If a system can imitate all others then it will imitate all primitive ones too that do have derivable complements. So, can we at least tell how these can be avoided?
Of course, one way to avoid them is to create again rich systems inside and then we simply repeat our original claim that these will lead to underivable complements.
A true insight would be if the avoidance of the derivable complements were achieved without this phony way, that is without referring to all derivabilities again. The perfect success would be a totally formal method that regards only the applications of the fix rich system without referring to any derivabilities at all. A half success would regard such formal method but also allow to look for individual derivabilities. As it turns out, only such half success is expectable.
This crucial method of underivable complements did come to light but quite late and distorted.
As the title says I will not reveal this crucial method yet. To give it more concrete form, first we make a detour.

Turing

After Gödel’s complicated proof of the Incompleteness of the $+ , \bullet$ arithmetic with induction axioms, within a few years it became clear to others that effectivity or as we call it, derivability is the main point. Most clearly Alan Turing realized this and he went as far as to say that effectivity is a new, mechanical or machine concept. To prove his point he had to create a computer model that was simple enough to accomplish two things:

Firstly, it was embeddable into logic and so the results applied to logical derivabilities directly.
Secondly, the mentioned imitation of all derivations in his fix system could exactly be shown.
This is what we could call inside universality.

The real vision of course was an outside universality too, that is the belief that all machine calculations, that is computers can be imitated in his system.
This was especially foggy because real computers still didn’t exist.
Yet the crucial element of the new computer vision was grasped. Namely, as the concept of “program”. It’s not really the machine with its fixed mechanical actions that do the calculations, rather the programs that tell the sequences of these basic actions. But the point is something deeper: John von Neumann and Turing realized that the programs are actually data too.
We can operate on the program steps as formal instructions just like they operate on numbers.
Turing saw that this means the inside imitability, the existence of rich or universal machine.
The strange thing is then, that this richness as universality is not just a potential feature of machines but also of programs and so data itself. All data, even the natural numbers.

The boring monotony of 1 , 2 , 3 , . . . hides this secret world. This vision later became called algorithmic complexity, which is just a fancy name for data being special through effectivity.
And Turing felt this quite clearly. The machine system that he created became known as the "Turing Machine" and for a while it remained in the background.
Real computers emerged and after the second world war, these things accelerated. But in the first major textbook on New Math, Kleene's Introduction To Metamathemetics, it was still the old fashioned number derivations that were regarded as the base of effectivity. Turing machines were mentioned as a merely more convincing way to justify the proofs. Kleene canonized effectivity, tied up the loose ends but still missed the crucial new result, we mentioned earlier. In fact he never embraced it even later. He continually returned to make his system of recursive partial functions more and more plausible. Amazingly, even in 1981 he presented a new system and he claimed that it is just as plausible as Turing's. And he was right because Turing machines as framework for effectivity are not plausible at all. That wasn't their goal.
The first immediate goal was applicability directly to logic. Namely to solve Hilbert's question about whether the logically valid statements that are theorems in every axiom system that is without axioms of a system are decidable. Here the word decidable is a translation of a German word that is not the same as deciding an individual statement. Rather it is what we call being totally derivable. But this is only clear with hindsight. So at that time it meant a method that could decide all the statements whether they are logical necessities or not. Hilbert made a parallel question about an other set of problems, namely the solvabilities of polynoms among the naturals.
The weird thing is that he could have made the same question about any axiom system, namely whether the theorems and the non theorems are a decidable set which actually means that the non theorems are derivable, because the theorems are obviously. The reason he didn't ask this was simply that he didn't see the point that derivability is behind both incompleteness that is undecidability of individual statements and undecidability of sets. So we shouldn't even use undecidability for sets rather the more precise claim of underivability of the complement, that is not total derivability. The obsession with total derivability as a primary concept stayed alive and still rules the field of effectivity. Turing was the first who clearly saw that derivability is the primary concept because the not necessarily termination of his machines were the corner stone of his results. But he went along with the old terminology and so exactly his new vision of effectivity as a potentially infinite process showed that finite decision is possible for those logical statements. By the way, the parallel result for polynom equations came much later.
But even Turing didn't see at that time why the total derivability is an even deeper fiasco. In fact, related to the yet unrevealed cause of underivability too. Exactly parallel with Turing, Kleene started to see too, that it is not decidability rather derivability that is the primary concept. In fact, he saw much deeper the mentioned fiasco but avoided to talk about it as a conflict and merely shifted the old decidability emphasis by introducing partial functions. The simple question why do we need functions then was never asked. Even though it is a trivial contradiction. Indeed, starting with functions before was logical because a function always has a value, it is decided. But then if we shift to partial functions that are not defined for all inputs then these are actually relations. So why cant we start with relations? We could but his tricks wouldn't work. Actually, Turing's machines do the same trick. They alter an input so calculate values. So it is understandable that Turing went along with the functional approach too.
But back to the first goal of Turing. Even though the actual proof of the undecidability of the Hilbert problem was delivered by his second goal, the internal universality and the underivable complement following from it, the real "proof" would only have been with showing that all effectivity is representable by Turing Machines. This is not only not plausible but impossible to show, yet seems to be true. It became simply overwhelmingly accepted due to computers.
But this doesn't make it plausible even today. So Kleene's resistance is admirable as was Gödel's.
A complete falsification of history happened later by some pinheads, portraying Gödel's and Kleene's reluctant submission to the undeniable fact that Turing did make a breakthrough and brought something totally new into math, as an open armed acceptance of the Turing Machines as effectivity. But apart from all this "who said what", Turing Machines are not a plausible framework of effectivity, period.

They could be called “Crawling Machines” because they don’t use memory registers. The infinite line is used as the memory and the data are one continuous flow of cells. Each cell can contain a single symbol from a keyboard. So the infinite line is one infinite text. To find, use and change something, we have to go or rather crawl symbol by symbol through everything.
Strangely, eventually we can still do whatever we want.
The changing is also cellular. We can write only a single symbol where we are at a moment.
The previous s symbol there is read before and helps to decide what the new s’ must be.
This s’ can be the same as s was, meaning no change but still better regarded as a writing.
We also have to decide which next cell to move after the writing, right or left.
If these writings and moves were made randomly we wouldn’t have a machine.
To be a machine actually means to make these choices mechanically that is by rules.
So the physical or material side is completely replaced by finding rules to make the writings and moves. We can even imagine a person using the keyboard but acting not freely rather by the rules. This person would see the keyboard and part of the infinite line where he will write using the keyboard. The keyboard has all the symbols plus two keys for moving right or left.
To achieve flexibility we should make our rules allow to depend on the data on the line.
Then strangely, the data line is actually a self altering set, using our fix rules.
The restriction of not using registers, dictates that we should only use the local symbols that we see around where we are doing the writing. We could even store a few, say ten we read last and decide by those rather than visually. The extreme is that we only use the single last read symbol where we are. The rest of the line outside this single cell could be totally hidden from our view.
Amazingly this is enough! Or rather, it is not enough but more read data doesn’t help, we need something else. Namely, finite choices that are not on the line rather could be called as our moods.
So, these internal m moods plus the last read s symbol will tell what we hit on the keyboard.
This means hitting a new s’ and right or left. But we decide something else from (m , s) too.
Namely our new m’ mood. So, the table of rules is (m , s) $\rightarrow$ (m’ , s’ , right / left) lines for every (m , s) combination. Having 10 moods and 20 symbols we have 200 lines.
For each possible combination of (m , s) after the arrow we find our next mood m’, the new symbol s’ replacing s and the move to right or left.
Thus, not only the line but we, the machine become a self altering set too.
The big difference is that the line is unbounded while the machine is bounded. It has only finite many moods. Yet the crucial fact is that this doesn’t mean that these moods are primitive in time.
They don’t have to go into fix cycles. We or the machine can be quite unpredictable.
This playing with the words “we” or Machine is hinting that the brain could be a computer, which of course no sane person should really believe. What is a fact is that:

The open space of the data line creates the open time of the machine mood sequences.
Starting from an infinite line with already stored symbols on it and with a given machine sitting in font of it, if we tell at what symbol on the line the machine should start and from which of its moods, then the alteration of the line and the machine is determined as two infinite sequences influencing each other.

In spite of this being a very restricted version of a computer, it is actually much broader than a real computer. And the “broader” is meant literally.
Indeed, the “humble” infinity of the line means an infinite memory.

This infinite memory means two things. Potential use for arbitrary wide calculations but also an above mentioned infinite initial condition. The first is a crucial necessity though not apparent at first. The second quite apparently contradicts the whole finite determination of machines.
Turing was fully aware of this contradiction and first was not quite sure what such infinite initial condition should mean. Since he was entertaining the idea of infinite decimals even for the calculations, real numbers as inputs could mean some interesting physical applications.

This never became realized only in a different version. So, gradually the pretty dull restriction became standard that we allow only a finite texts as start. That is infinity of empty space symbols towards the two ends of the line. But then you may ask why not go all the way and start from a blank line that is containing all spaces. And you are right saying this, in an even deeper level. Namely, the machine tables that seemingly only tell the transitions of moods and alterations of the initial symbols, can by a "side application" be used to create any finite initial symbol situation.
All we have to do, is add some moods to create these. So, finite initial data is totally avoidable!
And yet, finite initial data became the whole essence, the unavoidable point of Turing machines. To understand this contradiction is the first step to truly understand what von Neumann and Turing really discovered behind the concept of program.
At first glance the program of a "mood machine" should be its transition table.
Indeed, this is what tells how the moods change and the writings and moves must be carried out.
But as I said, the crucial point was that a program is itself data that can be manipulated, changed. The table is alterable too but only by us from the outside. So it's more like the hardware of the real machines. This shows an instant difference too. At real computers, faster processors and bigger memories are all that we can improve as hardware. A hardware change will not make the machine able to do something that it couldn't do before and we feel this plausible. Or maybe we don't and we are simply brainwashed to claim so. At any rate, here at the mood machines, the hardware, that is the bigger and bigger transition tables definitely feels like an infinite scale of complexities. But this illusion is false! Most importantly, it was created for a purpose! To show that a single machine with its fix transition table can imitate all of them. Namely, through the seemingly unnecessary finite inputs. These "unnecessary" inputs are the programs, the real essence. But a secondary importance emerged through this too, to allow inputs even for old fashioned data. This data dependant effectivity was the only way to truly express the universality of the mentioned fix mood machine. Indeed, otherwise we would again have to refer to the transition table simulations by those tables which are external. Instead, we can now prove our point by showing that the data alterations are simulated. But that's not all! This input data alteration vision became also the ground on which the mentioned true cause of underivability can be "recognized". I put this in quotation mark because this data view itself is a recognizability vision as opposed to our original derivability vision. But this doesn't mean that derivability is false. The non derivable complement is still the raw vision and recognizability is simply a twist towards this vision. This coincidence is that elevates Turing's awkward crawling machine from a mere original convenience to do his two goals, to a perfection.
Still, to regard it as a plausible way of effectivity is insane.
And clipping out Gödel quotations to prove this point does not make it true either.

In truth, there is no plausible method of universal effectivity. In either sense of the word universal. That is as a universal framework that includes intuitively at once all imaginable effectivities, and neither as universal ones in one framework by trivially imitating all others.
Both universalities are meant to be non plausible. But this will be clear only hundreds of years from now, when plausibility will be cleared itself.

The Cause Of Underivability, Part 2, Rice's Theorem

The class of real computer gigs became a class of New Math students that entered the scene. Through the internet. This is a system of evolution without natural selection. So the internet became the evolution of stupidity. The truth is still out there but it is the needle in the haystack.

Most shocking is that the final clear method of underivability, remained unemphasized.
In both camps, among the real mathematicians and among the new computer math preachers.
The concrete reason of this is that this new vision can be avoided and the formal result itself can be regarded merely as a quite easy consequence of Kleene's canonized results.
The fact that Kleene didn't realize it is an obvious proof of that a deep purification happened.
Kleene himself remained in total denial and his "offsprings" like Shoenfield's most extensive Mathematical Logic text book, still didn't even mention this new vision a decade later.

The Formalist evil that spread through New Math like cancer, is unstoppable.
Only a future didactical age will recognize the darkness we live in right now.
After these emotionally charged words it's time to tell what this new heuristic result was.

It is called Rice's Theorem. As I said, it can be formulated without even mentioning its heuristic consequence. Plus it's derivation was easy. These two facts resemble its true analogue, the most important Set Theory result, the Wellordering Theorem. So I reminisce about that for a second:

Back in Hungary when I fell in love with Set Theory in second grade High School, the only textbook available, was Kalmar's university text book.
It started with obligatory Stalin quotes but then covered most results in detail. In fact maybe too detailed. The Wellordering Theorem was five pages and I couldn't understand it at all.
I was attending a course for students who excelled in math, held at the university, together with students from a more famous math High School in Budapest. I remember that once I had Kalmar's book with me and one of those students talked about the Axiom Of Choice, so I approached him but he was just blubbering something that Kalmar's proof is unnecessarily long.
Later we became students in this same uni and his father, Janos Suranyi became our professor.
Unlike his annoying son, he was the most friendly and helpful person. But by that time I stopped asking questions.
A year later I was waiting for my US visa in Rome where I ordered Paul Cohen's book.
I used it as my first English course. The Wellordering Theorem was two lines in it.
Few months later I was working is the Stanford Math Library waiting for Cohen to return from England. That's where I found Shoenfield's Logic book and through it to the wall a few times.
I still had to have a copy of my own, so I ordered it in the Menlo Park book shop.
Somebody stole the one from the library and since they knew I always read it there, they asked me "If I knew what happened". I showed my copy and the receipt.
Cohen returned and I talked to him with my still very broken English.
The concept of Forcing "came" to me in High School and when I heard about Cohen's result I decided that I have to see him. The conversation was pretty strange and probably I insulted him a few times. The next day he looked for me and we had an even stranger chat. He apologized though he had no reason, except that he couldn't quite conceive my points which was probably my fault. Plausibility versus proof was already an issue for me and I mentioned the short proof of the Wellordering Theorem in his book as a negative example.
I returned to this theorem many times in the next twenty years and my final understanding only crystallized through these total restarts from scratch.

Unlike at the Wellordering Theorem where the other side, the Axiom Of Choice is at least spelled out, here with Rice's Theorem we don't have that other side yet.
In spite of this, the heuristic meaning is crystal clear.

Above, where I said that the underivable complement of the rich enough derivable set is not plausible, I didn't mention the worst problem. Namely, that this being "rich enough" is a totally artificial condition. I did mention that Turing made his machine so simple, in order to "see" the inside universality, that is the ability of some machine to imitate all others. But this universal machine is still a particular seemingly rare construction.

As computers became real, a new vision spread that they are all universal because they are all capable to compute everything. The PC shoppers look for faster processors and bigger memories and it's only the salesmen's limited math knowledge that stops them from grabbing the wondering customer by the am and say that "here we have a universal model".

Not surprisingly, they would be lying. Indeed, real machines have finite memory so imitating machines with bigger memories is impossible. The accepted "understanding" is of course that allowing arbitrary big external hard drive the machines are universal.

This still just masks the puzzle where the tour de force of Turing to demonstrate the existence of universal machine became trivial. Something is fishy here. The triviality of universality is a lie.

In both sense! So, real computers don't represent a silver platter that gives a plausible form of all effectivities and they don't show either why they can imitate even each other.

Instead of seeing this problem, the parrots just try to hypnotize everybody that the new computability age brought in a final solution. But it's actually much worse.

Repeating the words "Turing machine" in all subjects from pro or con artificial intelligence to the human brain and even aliens, became part of the parrot language echoing over the desert.

Of course this gerund sounding name "Turing" just rolls of the tongue. I wonder if poor Alan's father had been called Smith, would the parrots practice the same intellectual masturbation by repeating the "Smith Machine".

But back to real problems, the most obvious sign of trouble with underivable complements is that we can't give even one easy concrete example where this would be the case.

So we could simply repeat that just like we said, the modern trivialization of universality is false and Turing's blood sweat and tears are the only true source of non derivable complement?

But this leap from A to B is luckily false:

The first part is true, that is the trivialization of universality is false, but the second is false, that is we do have a better method and explanation of underivability than mere universality.

What was missing is the true connecting piece, Rice's Theorem. It gives a condition when the non derivable complement is not a rarity rather a must. And this condition is amazingly simple.

Even more amazingly, it is the original central concept, that von Neumann and Turing recognized.

The program! So the blind spot of this method was an even more amazing fact of history.

In our derivation systems we derive tuples from other tuples that obey some relations.

In short, our derivational systems are relational.

To be sure to get tuple sets with non derivable complements all we have to do is to use some part of our input tuples not merely as conditions of relations, rather as instructions of programs.

But crucially, we have to leave some parts as parameters open, that is practically we have to allow that these programs have themselves inputs. Then we can regard the program runnings without watching the actual formal program behaviors, rather only the parameters that is the inputs.

Namely, we can prescribe any easily decidable property about a finite set of accepted that is stop causing inputs. To collect such finite set is then effective because the individual runnings from the inputs to the stops are finite and the verification of the input property is too.

But now the complement set, that is those programs that don't possess such finite input sets are bound to be non effective! No machine can stop exactly from those programs as inputs.

This of course, makes only sense if the complement is meaningful. So the used inputs must represent all possible programs and so our system is actually a universal machine too.

So, universality is still a condition but the point now is that we get not one measly underivable complement rather a whole arsenal. Namely, for every recognizable, that is finitely verifiable condition of accepted inputs.

But as I said, amazingly the proof of this, is much simpler than it seems, because the only crucial point in the proof is that such recognizable finite feature of accepted inputs would obviously collect <u>all</u> the programs that run the same way. This operational completeness of the program set

is enough to derive that the complement is unrecognizable. In fact, the programs don't even have to be regarded as programs, rather as merely objects assigned to sets of objects.
This sounds bizarre. After all, such assignments of objects to sets could be pretty artificial.
But first of all, lets see what objects we are talking about. The old fashioned view is numbers or tuples of numbers. The new or linguistic vision is texts, using any symbols. Stupidly enough, the used word instead of text is "word" which feels much narrower than text. The point is not that we can use blank or space, rather that we have to allow arbitrary long "words", so they are texts.
Of course numbers themselves require already arbitrary long texts no matter how we code them.
The avoidance of tuples or rather merely regarding them as special texts is a big advantage of the new vision. Plus a program feels as a text, so the data being text already, seems like we are on the right track towards the big ideas of von Neumann and Turing. With numbers and tuples, derivation systems must be coded as numbers or tuples. But the previous argument that the programs are texts is only formal. The machines indeed alter texts but they themselves are still given as tables and it's not plausible how they can be translated to texts. So, assigning objects to systems that derive or alter or generate or recognize in short handle objects, is not obvious ever.
A crucial abstract point is that in the end, these systems will collect a set of objects, usually infinite, so assigning single objects that determine systems like programs do, also means that we were able to determine whole sets of objects by single objects. This set theoretical weirdness goes even weirder as we think about programs more. Namely, they are not unique for a system and a system is not unique for a collectable set. So actually, sets of objects must be determined not only by single objects but possibly by an infinity of alternative objects too. Almost like an absurd redundancy when even the objects should not be enough. To top it all, then we can repeat the crucial condition of such assigned objects have to satisfy, so that Rice's Theorem about them will carry the meanings of programs. Namely that we collect all alternative assignments.
Here we can ask, all alternative to what? To same systems or to same collected objects by systems? Seemingly this second one. So the systems themselves are not even involved directly.
This makes it even more important that we nail down a precise collection method to make more sense of all this. Recognizing is the most convenient.
In Turing machines this is easy to implement by merely accepting one of the moods as a start and an other or maybe more as stop. Then the recognizable inputs are finite texts from which starting at the left most non blank symbol, with the machine from its start mood, it will reach a stop.
This set is determined by the machine just like the derivable objects are determined by a derivation system. In a derivation system the freedom is to use the system continually. Here at recognizings the freedom is the singular lucky choice to start from an input that will cause stop.
Just as at derivation systems we can go systematically and derive all possible derivable objects in a sequence, here we should be able to generate all those inputs that lead to stops.
It's easy to say, lets try out all possible finite texts, but this seems impossible by two obstacles:
Firstly, how can we generate all possible finite texts? A dictionary contains all words in alphabetical order, so adding other symbols to the alphabet including the blank space with an agreed order in our new wider alphabet, we could have a "text dictionary" listing all texts.
This feels fishy and indeed, we made a mistake. Strangely it is not the adding of new symbols to the alphabet. The mistake is already there for words. We can not list all words. The only reason our human dictionaries do the job is that we don't have all possible words as meaningful. But even more crucially, the point is that we only have maximal long words. If it wasn't so then already the section A would be infinite. Indeed, we can make infinite many words starting with A and so we could never get to B. The solution is heuristically simple. We should organize our dictionary by increasing length! So instead of letter A, the first section is length 1. Within this we can be then alphabetical so list simply all single symbols in their order. This is finite many. Then we go to section 2, containing all two symbol texts again alphabetically. Then 3 and so on. The longer and longer sections will contain more and more texts but still finite many. So, all texts can be listed or generated. A machine can do this easily, so we can add this generating machine to any machine and thus come to the second challenge, to try these texts as inputs to test the real job of the machine, recognize some of them. Namely, by coming to a stop mood. And this creates the even deeper second obstacle! Indeed, if at an input our machine doesn't stop then it will never try

the longer inputs that may very well lead to stop. Luckily, we have infinite memory! So all we have to do is start processing an input and after a few steps start a next one. Do a few steps and return to the first. Then again start a third but return to the previous two. Then again start new ones but always return to all the earliers. This is called "dovetailing". With this heuristic trick, any set of inputs with a recognizable feature, can be "collected" as an actual generation.

We can collect programs systematically too that themselves collect certain inputs systematically.

We use a universal machine and enter programs of possible machines and possible inputs for those. The finite property of the accepted finite input set now simply means that after finite time we must come to it in our generation. So this is even more exact.

To see the point lets see a few examples: Whether a program will not collect the number 10 or will collect all primes is not verifiable in finite time. Indeed, the number 10 can pop up after any time and all primes can not pop up in finite time at all. The opposite of the first that is collecting the number 10 is recognizable by seeing it collected that is popping up. But the opposite of the second, that is not collecting all primes is again not verifiable in finite time because they all can pop up later. Collecting at least one or ten primes is of course recognizable again.

So, the set of programs that don't collect the number 10 or don't collect any primes or collect less than ten primes, are at once not derivable sets.

On the other hand, the programs that collect all or not all primes are complementing sets but neither is guaranteed to be recognizable in finite time, so neither of these complementing sets are guaranteed either to be non recognizable by Rice's Theorem.

The conventional proof of Rice's Theorem doesn't care about these programs and the two leveled recognizabilities. It hinges on the obvious and unemphasized fact that we collect all alternative variants of the programs because they run the same way. The programs are merely regarded as assignments to sets and we regard a set of them that is variant complete, so contains all objects assigned to same sets as an object already inside is assigned to. Then the complement of this complete set is obviously a variant complete set too but not obviously at all, it can not be a derivable set if the first was. Or to put it even better, for such variant complete pairs they can not be both derivable. Or, non of the two variant complete sets can be totally derivable.

The old fashioned, blind way to create assigned objects for derivable sets, is by creating these objects to the derivation systems, from their forms. The same goes for input recognizers.

These will never be operationally faithful, that is variant complete. This is why this seemingly much weaker condition that an assignment set is variant complete, is already enough.

But this is actually not true. We have to specify the assignments much more!

First of all, the assignment must be itself effective. This then means that actually we have a machine or system that collects objects and the assignment is simply a parameter of this machine.

For every fixed value of the chosen parameter the collected object set is effective and obviously dependant on the parameter value. But this value must be objectified. The easiest way to do this is if the parameter is also an input or part of the input of our object altering machine.

One direction of uniqueness by such assignment is then trivial from the method itself. One parameter value will mean a unique set of objects. The reverse is obviously not true. Different parameter values can bring about same collected set. We don't want to avoid this. Indeed, this is true about programs. We can alter or over complicate programs and they still will do the same.

To be universal, that is by altering the parameter being able to get all possible effective collections seems like a second necessary universality beside the universality of the used machine or system.

But this is still not enough. We need a third universality to obtain Rice's Theorem.

The parameter must be able to imitate other parametrizations. But here a strange reciprocity will come about. We can not expect that a fix machine or system could predict all such alternative parameters. Instead, we merely claim that there is a machine for every new parameter. On the other hand this is stricter than merely a derivation of the other parameter value. We can't stay in limbo and not get result. The existing machine must be an always terminating one and thus give a definite new parameter. The object alteration as the inner working of a machine is a perfect form for this but as we said the termination or stop is not a guaranteed fact for all inputs. This was the essence of the flexibility, namely allowing arbitrary long alteration sequences. This is also the essence of why non derivable complement seems plausible, namely because the recognizability by

stops is so complex and unpredictable. Finally this will be also the actual sure way to get the underivable complement from universality. So then how can we guarantee stop for parameters as inputs? We can't! We simply assume this as a condition. We can also see that this relates to total derivability because if an S set and its complement is also derivable then the two derivators as one machine means selection of all inputs to be in or out of the S set. So a yes or no result is ordered to the input definitely. Finite transformation or altering of course seems much more. We will come back to this whole problem because the total derivability itself will return many more times. In fact, we already saw that the final form of Rice's Theorem can be formulated through total derivability or rather as its impossibility too.
The more important point now is to visualize this whole parameter business. To call it as assignment or indexing or numbering is the usual but very bad way. Indeed, it doesn't express that a fix machine or system is involved. I will call it "listing". Unlike generating objects after each other in a line, the word list evokes the picture of being line by line. And indeed, here we list sets of objects each forming a line. To make it even cleaner, the set of the sets will be called as class. This is what we list, if we can. We'll come to some simple results about lists later.
So, the original proof of Rice's Theorem relied on program like listings that can be defined by the mentioned third universality of parameters. The final success of this abstract form of Rice Theorem was that it was provable that every underivable complement can be obtained from such variant complete "program like" set.
In spite of all these abstract results, in my opinion:
The real meaning of Rice's Theorem is about programs not listings! It says that:
Collecting programs by their partial, finite input successes, always gives underivable complement.
So, the programs that don't have such successful finite accepted input sets, are not recognizable.
We already saw that there are properties of inputs that give complementing program sets in limbo.
Neither input sets are recognizable in finite time and so Rice's Theorem "doesn't work".
There is an other sense in which it is not a sure method. Namely, it's not true that only these finite input properties can give underivable complements. Looking at the relational forms of derivation systems we are clueless how to get underivable complement. But allowing to look at program features in other ways, can also guarantee underivable complement. The most obvious is the universality, that is to be able to imitate other systems. But there are other similar ones. Similar in the sense that a certain self reference is involved.
After this grand view of Rice's Theorem we would expect that this will change everything for axiom systems too and we get better explanations of incompleteness besides underivability.
Unfortunately this is not the case and maybe it is an other reason why the grand view is ignored and self reference blubberings dominate the basic "understanding" of the whole field.
We don't know why Rice's Theorem in spite of being so grand is also so "useless".
Something is missing. So, we should just admit failure and return to the old arguments to get incompleteness. I will do that but this is a good time for two detours.
The first is a simple and very insightful proof of Rice's Theorem, using machines.
This will use heavily the existence of universal machines.
The second detour is showing how we can prove the existence of such for Turing machines.

Emptiness Acceptors, Proving Rice's Theorem

The start of our argument is quite strange. We pretend to defy all that I've been preaching up until now about the necessities of underivable complements as the whole point of everything.
In fact, we try to make a construction that will provide a perfect complement for any M machine.
The start is to continue the operation of M with an N second machine. By this we mean that if a t text as input of M is accepted by M that is M stops, then we let N use an s second text as input. We could even abbreviate this combined machine as t M → s N.
If we have a universal machine that can imitate M , N with m , n programs then with this universal machine our combined machine is the t m → s n program.
Suppose we have an E machine that uses programs as input and halts from all programs that are useless or rather empty because they accept no inputs at all. We don't claim that E halts only from these, that is recognizes the empty programs, merely that it accept all these too.
An obvious such E could be the machine that accepts everything. We could even exclude some concrete programs that we know are not empty. These are of course trivial examples of E.
A non trivial E would be such that it doesn't accept not merely some concrete programs but at least one full set of variants of an n program. The variants of this n are those n' that halt exactly from same inputs as n.
The usage of the same n as above is not accidental. We'll use this n variant completely not accepted program of E as our second machine. What will then t m → s n accomplish for particular t choices and regarded as a machine with the s input?
If t is accepted by m then it will be merely a variant of n and thus t m → s n is not accepted by E. If however t is not accepted by m that is running forever then the s inputs are not even tried for n and so t m → s n is an empty program and thus is accepted by E.
E of course has programs as inputs and the s input of these programs is not input any more.
On the other hand, we can regard part of a fix program as a parameter and regard the possible programs obtained by altering the part. So, we can regard now t as such and then it becomes an input of (t m → s n) E.
Most amazingly then this will exactly function as the complement of m.
This tricky switcheroo is not the false step that lead to a complement of the arbitrary m.
The really false assumption is E itself. Such Non Trivial Emptiness Acceptor can not exist.
Now we can use this fact to prove Rice's Theorem.
The E program examining machine had two variant complete subsets. Namely, all the empty ones among the accepted ones and all the n variants among the non accepted.
A drastic generalization of E would be an S super emptiness acceptor that:
Accepts all empty programs again.
There is at least one n program that it doesn't accept again. And finally:
All variants of a non accepted program are non accepted either.
This last requirement with the previous, obviously implies that all variants of n are not accepted, so this S is obviously an E too. Surprisingly, an "extra" feature is true too: Not just the non accepted inputs are variant complete but the accepted ones too. In short, both the stop set and the run set are variant complete. Indeed, variant completeness of one implies the other trivially.
The impossibility of E of course means that such S is impossible even more.
This implies at once that for two P , ¬P complementing program sets that both contain some programs, both can not be acceptable by machines, namely only that one can which does not contain the empty programs.
Indeed, if the one containing the empty programs were acceptable by a machine then that machine were a previous S super emptiness acceptor.
Finally, we can see that if we collect programs by the existence of certain finite sets of inputs collected by the programs themselves, then this collection of programs is variant complete and effective. So if we know that this collection is not total that is there are non accepted programs then this is actually a machine without complement. Also observe that the first theoretical part that we can not collect the emptiness programs is here trivial by the particular method. An empty program can not be collected because its finite collection sets wouldn't already exist.

Turing Again, A Table That Can Imitate All Tables

We are talking about Turing's transition tables that give for every (m , s) pair of m mood and read s symbol a triple (m' , s' , right/ left). That is, the new mood and symbol plus the move.
Before showing a universal table, I want to mention some interesting "unrelated" facts.
Writing texts, symbol by symbol brings up an older version of the computer keyboard, the typewriter. In fact, Turing recalled being fascinated by his mother's typewriter as a child.
Going backwards is a crucial advantage of computers, we can alter everything. How did we do without this, is quite unbelievable. Actually, typewriters did have back step, to use special erasers.
For Turing machines the back moves are much more fundamental. In fact, our universal machine will go back and forth to the program to decide every single writing and move.
But without allowing back moves, that is using an old fashioned typewriter but with an infinite carriage and replacing Turing's black box with a randomly typing monkey, we get something exciting too. I'm sure Turing knew about this set up called Borel's Monkey. The amazing fact is that if we wait long enough, the monkey will type down the full bible letter by letter. In fact, much more is true because we don't have to scan through the whole infinite half line to see where the bible starts again and again but fails after longer and longer lucky coincidences, to find a final full version. Instead, we can start by counting in a bible how many characters are in it exactly. Say a million. Then all we have to do is look at all the exactly million long consecutive sections of what the monkey has typed. Since we have infinite many such blocks, thus every possible such million long texts will be appearing by chance. So, one of these will have to be the full bible.
Now lets look for our bible. The universal table of Turing.
We already revealed that finite inputs will be the programs that imitate the tables and that we also have to regard the machines open to finite data inputs themselves. But this second feature is not crucial to define imitations. Namely, the originally mentioned more abstract alteration of even a fully infinite starting line can be imitated too. But with a fully infinite starting data line where can be our program? Well, simply between two special symbols say ! . If we cut out this segment, including the two ! symbols and reattach the line we have the arbitrary infinite data line.
The universal machine, using the segment between the two ! symbols can alter any initial line outside this segment exactly as any simple machine would do it reattached, that is without containing the program. The program of course depends on what machine we want to imitate.
The "exactly" above, needs some explanation too! It can't be truly exact already because the data line is broken so we have to cut out the program. But more importantly, the imitation will be achieved as follows: After we read a symbol outside the program segment, we crawl back to the program, find out what to do, crawl back and do the action that is the writing and the move. These are things that the machine we imitate, wouldn't do. More exactly, only the start and end would be there. Reading the symbol and doing the action. So the "exactly" means only that if we cut out these crawlings in time, just as we cut out the program section in space, then we would achieve the same alterations of the data line as the imitated machine does.
But this seems impossible, because from this method it's obvious that we couldn't crawl back to the place where we read the last symbol unless we place a marker there. Indeed, this second special symbol say ? thus seemingly intrudes the data line. Luckily, this ? can be put in place of the read symbol. When we return, we can replace the last read symbol. Do the action and then the new next symbol is read to its right or left. Then that will be replaced by ? .
So, if we cut out all those moments when ? is appearing, we wouldn't notice its existence at all.
The most obvious first question is how the hell can we remember the read symbol while we crawl back to the program segment, keep remembering this, plus then remember the action to be taken while we crawl back from the program. Obviously, we can not put these in our pocket because we don't have pockets. But we do have moods. Now the heuristic idea is this. We can multiply our moods in number by creating a new versions for each read symbol and actions to be taken. These parallel moods can operate the same way except when we want to make a distinction exactly by these differences. Now towards the second problem, how to turn a table into a single segment:
Obviously, all the (m , s) $\rightarrow$ (m' , s' , right / left) lines of our table must come after each other, each following an identification number that could be simply the repetition of a same

symbol. These addressing of the lines is not enough, we need corresponding addresses in place of the arrows using repeats of an other symbol. This will tell what the next m' is so with the read symbol can give a new line address. So a main cursor must tell where we are. This never disappears just moves in the program. The trick to identify addresses is to use secondary cursors that are replacing the repeated symbols used in the addresses. Starting by replacing the first one, then we can move these cursors ahead by actually rewriting the previous and replacing the next with the cursor. The identifying of the addresses is easy then. Indeed, the less many that is too short addresses are simply eliminated because the cursor in them gets to the end and can be deleted, while the too long ones can bee seen by having still address markers as next when the destination exactly finished. Of course this method works by going or rather crawling through the whole program as many times as the number of an address. Plus the re-settings of cursors requires additional full crawlings.

A third special problem is to take care of the two sides of the line. The program segment must allow to go over it and continue on the other side as it were continuous. Of course this comes when a read symbol is the ! separator ad we can't allow to replace this by ? rather do that on the other side at the first symbol after ! .

The p program including its two ! end markers is a text inserted into any initial data line, so that an M machine or table operating on this line without p will be exactly imitated by our U universal machine that operates on the line with p. The only ambiguity is where the M should start on the line and internally. So, we can pick a mood of M as start, and agree to start from the next cell after the right ! marker and make our U accordingly. Then the infinite run of M is perfectly imitated by U with the mentioned meaning of cutting out space and time.

But the real goal of U was much more humble, namely to imitate merely the alterations of finite inputs or even more humble as imitation of stoppings from the same inputs. These are of course at once included if we specify where the finite input is regarded and what moods are regarded as stops. So U (p) imitating M at once means U (p , t) imitating M (t) where t is any finite text. To make this form look even more meaningful, we can agree that t is actually assumed to lie next to the p program and we already agreed that M starts from there too.

If p is the program of M then U (p , t) stops from a t exactly if M does.

In fact, the imitation is better because not only the stopping but the whole alteration of the line will be the same. This brings in a formal problem nothing to do with the promised twist.

In classical math this bracketed notation we used above after M and U was accepted in two very different sense. For states that is properties or relations like P (x) , R (x , y) and also for functions like f (x) , f (x , y , . .). The big difference is that the first is a yes or no claim the second is an object. This drastic difference with same form caused no problem because the context always told which is the case. But not accidentally, here with machines we came to a point where this becomes apparent and also brings out two hidden layers in the confusing form. We could simply say, what's the big deal? M (t) could mean both the stopping from t as claim, or the altered text from t at the stop, so we have to make a distinction in notation. And indeed we do this by accepting the second that is the object meaning as basic and use $M(t)\downarrow$ as the claim for stopping. In fact for the negative that is $\neg M(t)\downarrow$, we use as abbreviation $M(t)\uparrow$.

The best reading of $M(t)\downarrow$ is: "t is an M-stop input" while at $M(t)\uparrow$: "t is an M-run input".

So what are the two deeper problems? The first is that the chosen M (t) object meaning as the altered text from t is actually dependant on the secondary $M(t)\downarrow$. Without stop there is no value of M (t), it doesn't exist. This bitter truth that is self evident for machines, came through only after a lot of pain among old fashioned functions, when finally Kleene accepted partial functions, not defined for all inputs. There is even a new notational equality arising from this hidden object existence assumption too. If we write M (t) = t' this should also include the existence that is $M(t)\downarrow$ but more interestingly $\forall\ t\ [\ M(t) = N(t)\ ]$ should not mean that for all t both machines stop rather that they stop from the same ones and have same values.

Kleene also accepted this meaning as the equality of partial functions. In fact, this $f \cong g$ is even read as f and g "return the same".

The second deeper layer is that actually $M(t)$ is a false form altogether. Indeed, M operates on an infinite T text and step by step. So $M < T >$ should be a single alteration of this infinite T. $M(T)\downarrow$ could be the stop from T and $\{ M < T > \}$ the sequence of line alterations up to stop and $M(T)$ the final line. All these can be used with $T = t$ finite initial text, that is blank towards infinities. After these formal reflections we should return to the real problem with $U(p, t)$.
We can meaningfully write now that $\forall M \; \exists p \; \forall t \; [M(t) = U(p, t)]$ and yet we couldn't apply our final trick to show that a machine exists without a complementing one.
Obviously this machine would be obtained from U and the problem is not the difficulty of the crucial trick that I intentionally not reveal yet, rather that it can not be applied. That trick will be introduced with sufficient fanfare as a whole section claiming to be an alternative birth of New Math. But this problem we have now has nothing to do with that. This dirt in the machines is clogging the wheels and this is much deeper than simply a glitch that we could ignore and jump for joy as practically all "educators" of the field do. In fact, this "cover up" goes back to Turing's original article. The surface of the problem is simply that in $U(p, t)$ we have mixed messages.
The p contains the extra symbols we used to operate U. The time marker ? is the most alien and intrusive. An easy solution would be to say okay lets code the p program into the alphabet of the M machines we imitate. Even the simplest binary $0, 1$ alphabet is enough to code arbitrary many symbols, in fact we could assume that any coding of a p starts with a 1 and doesn't use more consecutive 0-s as say ten. Then, since we are imitating finite inputs, we can assume all blanks towards infinities. So, the full data of our new U' binary universal machine could start from the leftmost non blank that is 1 symbol. That's where the coded program $c(p)$ starts and we can use ten 0-s to finish it and separate the t data lying to its right.
These binary $c(p) + \text{ten } 0 + t$ inputs thus must be altered as follows:
Use the d decoding procedure for the segment from the leftmost 1 up to the first ten 0-s, that is for $c(p)$ and thus $d(c(p)) = p$ is recreated. Then apply this decoded p as program with our earlier described U universal machine from the segment following the ten 0-s, that is t.
Of course, this binary machine U' could only produce the alterations that are binary and don't have the special symbols. So: $\{ U' < c(p), t > \} \subset \{ U < d(c(p)), t > \} = \{ U <p, t> \}$.
The $\subset$ symbol means containment, namely ignoring the space section between ! and the times when ? appears. To find this U' from U seems hard and nobody gave it ever effectively.
Okay, so can we at least prove that such U' exists? Well, if we assume that all effective alterations of some texts in an alphabet can be expressed with tables using that alphabet, then U' must exist. Turing obviously believed in this, in fact that was the foundation of his claim that all effectivities are expressible in his machines, so he solved Hilbert's problem.
And yet he didn't use this argument, he created a concrete U' with a new trick to go from U.
He used every second cell for data which is an expansion of the memory. In effect this means a compression of the data. So we don't code the programs rather the data. This way, we accept the special symbols as part of an alphabet and we use a decoding of the real texts, the inputs, to avoid conflict with the special symbols. Again, the minimum for codings is two symbols, so we can use $0, 1$ for this. Actually, these can be part of the special symbols, that is can be used in p if they don't interfere with the data alteration outside p on the data line. So this way the special symbols are all are all we need for U' too. The containment of U' in U now is:
$\{ U' < p, c(t) > \} \subset \{ U < p, d(c(t)) > \} = \{ U <p, t> \}$.
Here p and t use the same special alphabet. So the usage of U' is not to create same alphabet like above. Rather oppositely, to transform this special alphabet for t, using $c(t)$, into the non interfering $\{0, 1\}$ subset of the special alphabet and thus being able to use p correctly.
The real point of course is that now this U' can be given concretely using any reasonable c.
The fact that we had to use the special alphabet would suggest that the reduction of this could be vital, but not really. Once we have any fix alphabet with a U' universal machine, we can code all texts to binary and add a lot of new moods to U' to create a binary U''. Shannon showed this trick and also an opposite one, that can create a merely two mood universal machine by using much more symbols than U'. He also asked the question how far the two numbers, the moods and the symbols could be reduced simultaneously, which is usually measured by their product. This chase for "simpler" universal machines relates to the mentioned cover up. Indeed, it gives the

false impression as if a large number of moods and symbols would imply the existence of a universal machine trivially. The truth is that Turing machines are not a natural construction so they shouldn't provoke any plausibilities and they don't. Only pretentious learnt reflexes. Making up big claims based on personal beliefs and then using Turing machines as a support either by simply repeating it as parrots or going just one level above as monkeys that use their hands too. These monkeys are actually worse than the parrots because they play with some details of the Turing machines. So they seem to go into it and deceive the naïve bystanders, in our case the clueless readers who are lazy to start math at the bottom and rather pay a fortune for these "enlightening" books that then can end up on their bookshelves as conversation pieces. Gödel Bach Escher, Emperor's New Mind are two perfect examples. This second is even more sad, coming from a real mathematician Roger Penrose. I will show without any shadow of a doubt that Penrose has no idea about what he is using so passionately. His beliefs about the mind and machines are very close to mine, so it's even more important to tell these. These are not mistakes or oversimplifications but intellectual and moral flaws. He is a liar and he has enough knowledge and gut instinct to see that he is lying.

But first lets go back to our unsuccessful idea where we code the programs not the texts. As I said: We could rely on the external universality of Turing machines. Turing did rely on this assumption for his claim that he solved Hilbert's problem but he didn't rely on this to finish up the existence of the internal universality. Instead he went the other way. He didn't directly compress the data rather used every second alternative cells for it, so expanded the memory. But why?

To understand this, we have to see an other not quite cleared part of his article beside "proving" Hilbert's original question A crystal clear reference is made about Gödel's result. Turing explains in one perfect sentence that if completeness would stand then the non theorems were derivable as a set. So this in fact means that his article re-proves Gödel's Incompleteness Theorem.

But this can only stand as a true derivation if the inner universality is established as a fully inside business without referring to the bigger claim of his machines being universal externally.

So actually his article solved not just Hilbert ill formed and thus unprovable question, rather found a better new proof of Incompleteness with effectivity. Of course, the belief that his machines are universal externally, that is can imitate all mechanical decision methods is also true. So Hilbert's problem is established too. But this is not only a wider field but a deeper too.

A less deep bizarre fact is that Turing used not strictly Turing machines rather ones with built in sub machines. A hard article to follow but crucial for at least to see, what initiated all the truth and bullshit about it. A mini version of the experience reading Newton's Principia.

We could wonder why smart people don't write notes to these big original works to clarify and deepen them? Because smart people are Formalists. They don't give a rat's ass about enlightening others. They rather laugh about the junk vomited out by those who preach falsely.

Which brings us to Penrose. His description of the universal Turing machine is a scandal.

It is not the above mentioned cover up of not to assume external universality. That cover up makes an unexplained turn away from the seemingly simpler program coding to data coding but still contains the previous tricks to show universality. So, the problem of linearizing a machine that imitates other's operation on an input is faced. Penrose is merely linearizing the tables of the individual machines! Which of course is trivial. Each is merely the skeleton of a program where the whole point is the additional fix special symbols. They make it possible that an external table can run back and forth to the outside line and alter the cells as the skeleton table would. Penrose doesn't show this whole problem of two tables and the crucial necessity of a marker in the data and more in the program itself. Now, we might think that he simply uses a stronger version of the external universality. Then it would be an acceptable simpler explanation. But this is false. He lists the machine codes and then just assumes that a complicated enough machine will exactly behave by regarding an initial segment as a program and execute it on the continuation. This is not an assumption of universality of the tables. If we have a fix mechanical procedure then it can be implemented by some table. He doesn't show a procedure, merely describes an infinite wish list.

From Underivability To Incompleteness, First Approach: Underivability Of Truths

We already revealed that the heuristic idea is to get incompleteness from underivable complement. But there are four different kind of arguments for the details to suit axiom systems.
The first applies to any axiom system that uses $+$, $\bullet$ and claims only axioms that are true among our real naturals. A completeness, that is every B statement being decidable means that exactly the true statements are derivable. Indeed, since reality obeys Logic, if our axioms are true we can only derive true statements. If in addition, our axioms are derivable too by some rules then the continuation with Logic means that the theorems are a derivable set. So to prove the impossibility of a complete system, enough to show that the true statements is a non derivable set.
The already revealed importance of non derivable complements would suggest that the true statements is such set and so its complement, the false ones were the derivable. But this is not true, they are non derivable either.
Both being non derivable? How the hell can we prove that?
First lets see the obvious, that if the falsities were a derivable set then the truths were too.
Indeed, suppose a D derivation system derives all falsities. Applying a formal negations after D that is $D + \neg$ would derive all the truths.
In spite of our claim that all the truths are non derivable, we can still find sets of truths that are.
In fact, Gödel's heuristic tuple decodings proved that the derivation systems can be replaced by explicit formulas using $+$, $\bullet$. So if there is an S set of derivable numbers with $\neg S$ being non derivable, then we can also find a $P(x)$ explicit formula that represents this.
That is : $n \in S \leftrightarrow P(n)$ is true. Thus of course:
$n \notin S \leftrightarrow P(n)$ is false or: $n \in \neg S \leftrightarrow \neg P(n)$ is true, that is $\neg P$ represents $\neg S$.
But $\neg S$ is a non derivable set so the $\neg P(n)$ statements that are true are not derivable either.
Now, we might jump to the conclusion that then obviously the wider set of all truths are not derivable either because they contain these special truths and these would have to be derivable. And next we can jump to a second conclusion that then for any derivable set of axioms the logical consequences give a derivable set of theorems which thus can not be identical to the full set of truths. So we must have undecidable statement and so our axiom system is not complete.
Both of our conlclusions were false.
The first was a true consequence but falsely over simplified.
Derivability is not being covered by the derivable cases of a system rather exactly being only those cases. So, just because the full set of truths contains the $\neg P(n)$ truths, it doesn't mean that the wider set being derivable would imply the smaller set being too. But here it does because this form itself is a recognizability among the true statements. So a recognizing of the true statements among the statements or expressions in general, could be followed by a recognition of these forms and thus indeed all $\neg P(n)$ truths were derivable.
The second conclusion was out right false in its first part that is claiming undecidable statement.
This is obvious by regarding the axioms $\{A , \neg A\}$. It is contradictory, so anything is derivable from it including all the truths, so there is no undecidable statement. Incompleteness of course stands because complete means all statements decided only one way but this trivial contradictory incompleteness is not our goal. But proper incompleteness, that is undecidable statement can easily be achieved by assuming that the axioms are not only derivable but true too. Indeed, then their logical consequences are all true too, so the theorems can only be a subset of the truths. The full set of truths being not derivable then means that the theorems are a proper subset of the truths. So some truths are not derivable. Their negative of course is not derivable either because the axioms were all true. So these unprovable truths are undecidable statements.
The contradictory axioms as counter example raises the question whether instead of this easy way, assuming truth of the axioms, the opposite condition of contradiction, that is consistency, could be sufficient to achieve undecidable statement.
That's our aim but the next second level is still not this.

Second Approach: True Axioms

Next we still look at the naturals with $+$, $\bullet$ but start from our simplest axioms.
This includes all the consecutiveness axioms but using " +1 = " instead of $\vartriangleleft$.
So we have the infinite many name axioms and the three ordering axioms of $\vartriangleleft$.
Plus the two pairs of derivational axioms for $+$ and $\bullet$.
As we explained, this system is very "raw" because it is trivially incomplete. Yet on the other hand, it is not that raw because it is sufficient to represent every derivable S set, which now means existing P (x) property that: $n \in S \leftrightarrow$ P (n) is theorem.
By the way, as we mentioned, already only boundedly universal P (x) properties can do these representations. This is plausible because the used variables in our derivation systems are also all existential and only the target variables are universal. The final plausibility is of course Gödel's heuristic tuple decoding idea that allows explicit formulas.
The existence of derivable S set which is not totally derivable, that is $\neg S$ being non derivable means that for the P (x) representing S, we have: $n \in \neg S \leftrightarrow$ P (n) not theorem.
This also means that the set of these non theorems is not derivable just as the $\neg S$ set itself.
Indeed, if it were then we could recognize the n numbers in them and collect them as $\neg S$.
Also, since P (x) is a recognizable form, thus the set of <u>all</u> non theorems is a non derivable set either. Indeed, otherwise among them we could derive the P (n) formed non theorems.
Now, suppose we want to accept new axioms, that is extend the above "raw" system.
If the representation of the derivable sets remains, we instantly get the same result that the non theorems are a non derivable set. Now comes the heuristic jump to incompleteness:
If this system were complete then a trivial way of deriving the non theorems were simply deriving the theorems by the Axioms + Logic and then applying a negation formally.
Beside the assumed representability of the derivable sets, we made a hidden one too. Namely, that our new axioms are a derivable set, that is the axioms are given by some rules or recognizability.
Indeed, only this way will Axioms + Logic be a derivability chain.
This derivability or recognizability of the axioms is of course always trivial and that's exactly why it wasn't recognized for so long as an important issue. But now the temporary bad news:
Accepting new axioms can easily ruin the representability of the derivable sets. The good news is:
Accepting only new axioms that are all true among the naturals will not ruin this.
Indeed, since the true bounded universalities are all derivable in the raw system already, they remain. And a false one can not become theorem in the extended system either because if it were consequence of true new axioms logically, it had to be true.
So our hidden agenda to start with the raw system is coming to its perfect use.
Lets repeat first that old agenda:
Gödel's original proof regarded the induction axioms. This was understandable. They were believed to be the magic wand that can prove everything. In classical number theory they are still the magic wand and we have to accept their heuristic power. But they are not complete, they can not decide all statements. This was the bombshell of Gödel. Of course, as it turned out, this has nothing to do with the actual induction axioms. It's not that they are not heuristic enough. In fact, the even bigger bombshell is that the real reason of incompleteness lies outside, in the seemingly so raw yet not so raw axioms. They are obviously incomplete but as we showed, the non theorems in them are a non derivable set. What's more, in any extension that is true among the naturals, the same remains. So, applying it to the induction axioms, if they are true, which is very plausible, then the non theorems remain to be a non derivable set. So this system can not be complete either.
This is the power of underivability! It showed to be the real reason of incompleteness in two sense. Firstly, it showed that the induction axioms are not the cause, they are not weak or faulty.
Secondly, it inherited from the raw system to the full induction system and silently turned the trivial incompleteness of the raw system into the absolutely non trivial incompleteness of the full.

Third Approach: Total Derivabilities

But we still haven't answered our earlier question whether consistency, the minimal requirement of a proper incompleteness that is undecidability, could be enough instead of truth.
Indeed, sometimes we don't regard merely the naturals. In fact we might not have any fix model that we describe, rather common features of different models. So then our axiom system intends to be a theory rather than a description of a single reality. Arithmetic can still be part of such system and then if it grasps multiplication, it can't be a complete theory. Of course from most theories we don't even expect completeness. At any rate:
We already revealed that some extensions of our raw system can ruin the representation of derivable sets so we might think that this means the end of the story. Luckily, there is a layer under underivability that might be regarded as an other, even better cause of underivability itself.
The formal idea is so simple! Why should we care about representing all derivabilities when our goal is underivable complement? We should only care about representing all totally derivable sets. Then some indirect reasoning can show that a full total derivability is impossible, so there has to be underivable complement. The real extra beauty is that this way the existence of underivable complement will not be assumed externally. Though strangely, in this third approach it is still external, above the system. Then the fourth approach will be the truly inside self reference view.
The old rule that made all this work is again very simple. Less can be more if it's better.
For representing only the totally derivable sets, this better is quite natural:
Indeed, a simple representation means a $Q(x)$ to represent S and an $R(x)$ to represent $\neg S$.
A perfect representation would be a $P(x)$ for S, so that $\neg P(x)$ is for $\neg S$.
Amazingly, this P is quite easy to obtain from Q and R if we accept $<$ in our raw system with some added simple axioms about it.
With such perfect representation of the totally derivable sets, everything becomes perfect.
First of all, the derivabilities of these $P(n)$, $\neg P(n)$ case pairs can not change by adding new axioms to our system unless it becomes inconsistent. Exactly one of them is theorem already.
So, all consistent extensions of our system will also perfectly represent the totally derivable sets.
The goal is to show that with a derivable set of new axioms the non theorems are a non derivable set, but to be even more open we can just say that with any new consistent axioms the set of theorems or the set of non theorems must be non derivable. In fact we feel that this "or" was not an "either or", that is some axioms can make both the theorems and non theorems non derivable.
But with these plausibilities we have to be careful and I make a short detour to make this point:
We already know that our raw system has $P(x)$ that represents non totally derivable set.
Now, we might think that this $P(x)$ remains the same. We might even think that all such properties of our raw system that represented non totally derivable sets should remain the same.
The "logic" is this:
Adding a derivable set of new axioms, for any property where the non theorem cases were a non derivable set should not become derivable. On the other hand, if we add a non derivable set of axioms then the theorems will become a non derivable set. But beside not being precise, this plausibility is false by the following "counter logic": Our raw system is itself built from axioms that created the assumed absolute system of total and non total derivable representations. So that system is not absolute. Tinkering with axioms can change the derivabilities not just individually but as sets. So there is a healthy plausibility behind the absoluteness of derivable versus not derivable sets but this absoluteness is not true in the field of altering derivation systems, which includes tinkering with axioms. To tell exactly in what field of actions must derivable and underivable sets remain is a deeper question. The incompleteness or rather undecidability claims only rely on that the existence of underivable complements must remain. By the way, this assumes consistency too and this shows the fault of the previous "plausibility" too. Minute change can cause inconsistency which of course changes all derivabilities, including the derivable set representations.
So back to our more general claim:
A consistent extension of our raw system can not have both derivable sets of theorems and non theorems. In even shorter form, these two sets can not be totally derivable.

An Alternative Birth Of New Math, Infinities

The fact that these totally derivable sets came in again is a weird reoccurring fact. They already tried to rule Rice's Theorem and I promised to go deeper. Now we go deep and strangely still not finish this total derivability issue. In a sense only start digging into it.
But going deep is meant fully, so actually we restart New Math from scratch.
I claimed the birth to be Beltrami's recognition that models obey Logic. But started from Euclid's parallelity axiom, which thus could be called the pregnancy. I also smuggled in the concept of sets that provided the new realities as structures or models to break away from the earlier physical realities. So the alternative start is to emphasize Set Theory and its creator Cantor. If this is the birth then strangely the pregnancy again goes back to Euclid.
Aside parallelity there was an other "parallel" problem namely with distances.
In physical reality, that is measuring them for cabinet makers or builders, the measuring tape is the solution. The identity of two distances is then perfectly establishable by rolling out the tape for both and having the same length on that. The actual verification of this would be marking the tape. We don't do that, rather read the already existing marks on it. But what if the distance is not exact marked point, rather in between. We simply accept the errors and use the next marked value under or above. The Greeks asked the theoretical question whether with infinitely finer and finer markings could we always measure exactly. First they thought "yes", but then they realized "no".
The finer and finer markings are of course fractions of a unit of the tape, or so called "rational" numbers of the tape unit. So the negative result is that there are irrational lengths.
First this whole problem seems insane because if there are irrational distances then they are already on the tape too. So the whole problem should be stated whether the fractions on a tape would fill up the tape or will leave holes. The fact that this formulation of the problem didn't survive is again an example of the already existing ancient Formalism. Indeed, the success to show that there are holes came not by examining the tape that is one single line, rather going out into the plane. Namely, the d diagonal of a square can not be a fraction, if the s side of the square is the unit. Indeed, putting an other square on d, its area is four identical pieces from the half of the original square. So the diagonal square is twice the area of the original one.
In short: $d^2 = 2s^2$. This is also a consequence of the Pythagoras Theorem used for the half square triangle, with d and two s sides. The claim that with s as unit, d can not be a fraction, means algebraically that $\left(\frac{n}{m}\right)^2 \neq 2$ or $n^2 \neq 2m^2$ for any m, n whole numbers.
The Unique Prime Factorization of numbers implies that both m and n have fix predetermined number of 2 occurrences too. It can be zero too for odd numbers obviously. This unique factorization then stands for the squares of m and n too and thus, these must have exactly twice the 2 occurrences than it was in m and n. So this occurrence must be even in both square numbers. But in $2m^2$ the 2 occurrence then must be one more and thus odd.
So the two sides can not be equal for any m and n whole numbers.
What a weird thing just happened! We had to go from the line to the plane and then finally it all boiled down to natural numbers and their multiplications which seems to be strictly a business on the line, in fact only among the whole multiples of the unit. That's reality beyond sets and logic!
Multiplication is not a strictly whole business even for the naturals. This is apparent from an other fact too. Namely, the exchangeability of the multiplication order which was not following from the derivational rules without induction and is also not plausible on the line, becomes obvious in the plane as area of rectangles.
The irrationality of distances became a well accepted fact and slowly even the elementary educational vision reflected this. The crucial silver platter that delivers this, is the decimal system or Arabic numerals. Of course it's not the number 10 as base, rather the digital representation as power sums is the point.
There are three levels or consequences of this silver platter. The basic representation of the natural numbers. The digital calculations especially the division process. And finally that connects with this is the infinite decimals as points of the line.
Already at the first level, the use of zero, jumps out as a most obvious side effect of this system.

These Arabic numbers came from India where they used zero thousands of years ago and calculated with billions. An early pope declared the zero the devil's work but the awkwardness of the roman numerals had to go by simple practical reasons. Already before, when the non believing soldiers ruled Europe, the collection of taxes was a nightmare and now when slavery started to disappear, peasants had to count. So it became obvious how superior these new numbers are.
When the elementary school teacher explains that: $365 = 3 \bullet 100 + 6 \bullet 10 + 5 \bullet 1$, then this is merely a formalization of something the kids already know. Why this composition from increasing powers is so good is still a hidden mystery then. The digital calculations later are again merely some tricks that are nowadays even avoided due to the use of calculators. And it really doesn't matter much. The crucial next step is what the pope didn't know yet. The real magic.
This magic is the decimal point. This tells how an infinite decimal is locating a point on the line:

$$365.74903264\ldots = 3 \bullet 100 + 6 \bullet 10 + 5 \bullet 1 + 7\,\frac{1}{10} + 4\,\frac{1}{100} + 9\,\frac{1}{1000} + \ldots$$

The arbitrariness of composing the whole numbers from powers, here becomes the natural equal and repeated divisions of a unit. This at once shows that a finite length is a sum of infinite many smaller and smaller intervals. Then using the same for time intervals, we at once see that a point in time can also be approached by infinite many points of time. There are infinite many moments approaching the "now". Like a minute ago, a second ago, tenth of a second ago and so on.
So "infinite many times" doesn't mean "forever" and something being this or that way infinite many times doesn't mean being that way forever either. So, Achilles being nice and giving a head start to a slower runner will also simply mean that infinite many times he will be behind the slower runner but these instances don't mean forever. They approach a point in time, namely when he surpasses the slower runner. Achilles paradox is totally resolved hail the silver platters.
This didn't resolve other paradoxes of Zeno or Parmenides concerning deeper problems of points that we simply ignore today. But the digital division process using the remainders, tells at once that a fraction is always an infinite decimal that has some s starting segment and then a repeating r segment. Indeed we can only have finite many remainders, so the division process must return to an earlier remainder. Using calculators, this is merely experienced as a fact, that we have no explanation for. So learning the division process alongside calculators can be even more awakening. Most importantly, what the division of naturals, that is the decimal form of fractions show evidently, is the existence of irrationals. Without going out to the plane. The infinite non periodic decimals are obvious "holes" on the line. They are not fractions.
Two even more specific ways of creating such non periodic irrationals is to use random digits or quite oppositely, tell them by rules that definitely avoid to be periodic. Deep problems can be raised by these two and yet these were not raised before new math, even by mathematicians.
Instead, a different direction of generalizing the rationals was coming from solving equations.
Solving the first order equation $m\,x = n$ with m , n whole numbers gives the fractions but solving higher order equations using only wholes, we get irrational roots most of the time.
The natural question was whether all possible numbers are solutions of some whole equations, or as it is called are "algebraic" numbers. Everybody guessed this shouldn't be true but it was quite hard to finally exhibit a method that creates infinite decimals that can definitely not be algebraic, that is can't be roots of any whole equation. The genius Gauss didn't find this method, so it shows that it wasn't that trivial. The important part of course is that there was a sword to cut through this Gordian Knot and Gauss didn't find that either. It's hard to guess how he would have responded if he had been alive when Cantor found this sword.
All this prelude was important because there is a misconception about how Cantor created our paradise, Set Theory from nothing and relating to nothing at the start. This is not true! He solved again the old Greek problem of irrational numbers, he solved again the quite recently established existence of non algebraic numbers at his time. And he proved the existence of numbers that would defy an infinity of similar possible constructions.
In fact with a stretch of imagination he proved that even the above raised natural dilemma is solved partially. There have to be irrational decimals, that are not completely ruled.

The partialness means that this still leaves the random ones open. Indeed, being random is more than being not completely ruled, namely not having ruled part at all. So the random numbers are a special subset of the non completely ruled ones and thus their existence doesn't follow from the existence of numbers that are not completely ruled.
We'll come back to this whole side line in the next section.
Without even revealing Cantor's genius idea, these facts make us wonder how the hell it was possible that some opposed him. And it was not only the devious Kronecker. Exactness of proofs was already a common feature, so quite obviously the opponents simply didn't accept his method as a proof. The problem was not that he proved facts that were already known. Mathematicians re prove thing all the time and for him to find a combined and simpler method should have been enough to gain recognition. This is a very crucial departure from merely blubbering or smiling over how things went. Most importantly, it is a departure from the bullshit that "his ideas were too revolutionary". He created proofs! Being revolutionary is an opinion, a proof being faulty is a fact. The morons of today who praise Cantor but don't think about these points are exactly the ones who wouldn't accept him if he lived today. Most importantly, if this paradox could happen, it will happen again. Yes it was a paradox and no matter how much I hate Kronecker I will never regard this paradox as a psychological thriller either. The acceptability of a proof is not as absolute as we think. And the formalization of Logic did not stop this paradox to reoccur.
Okay, enough of this spewing let's see how Cantor showed the existence of irrational, non algebraic numbers and the rest.
The start is something that actually goes against the crucial final idea, which is to accept that the infinity of all possible decimals is much bigger than the fractions or algebraic numbers or other ruled decimals. The counter idea is that infinites are all the same. They swallow up everything.
A surprising historical fact starts with how Galileo measured the falling distances in consecutive seconds of a dropped object. He realized that these are: $h \ , \ 3h \ , \ 5h \ , \ . \ . \ .$
He was amazed by this and thought that some deep reason lies behind. The simple reason is that: $h + 3h + 5h + \ . \ . \ . + (2n - 1)h = n^2 h$. So the total fallen distance as $t^2 h$ is not only simpler but it at once gives the distance for non whole times like say a half second. We might deduce that he didn't have the mathematical talent that Kepler or later Newton even more possessed.
Indeed, he missed the good math that helps physics but amazingly, he wondered into much deeper math at the same time. The fact that the odds are a half sequence of the full set of naturals and yet are themselves a full infinity fascinated him. But it wasn't his time yet.
To splinter the naturals even more, was discovered later by others:

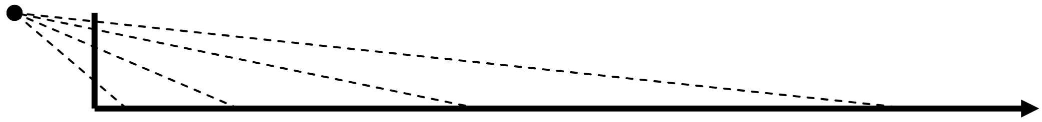
1	2	4	7	11	.	.	.	.	.
	3	5	8	12	.	.	.	.	.
		6	9	13	.	.	.	.	.
			10	14	.	.	.	.	.
				15	.	.	.	.	.

We only used every natural once and yet created an infinite sequence of sequences.
We also showed earlier how Gödel created codes for all the tuples of naturals, though there the reverse coding and the fact that he only needed multiplication was the real point.
To sequence merely the pairs of naturals like the fractions is much easier by simply listing them as groups using the increasingly bigger and bigger numbers as numerators or denominators:

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

If we visualize these same fractions as a dense point set "filling up" the whole positive half line, it makes us even more believe that all infinites should be the same.

A totally different argument for this same belief is that we can project an arbitrary small interval onto an arbitrary big one. The points as elements of these sets magically correspond to each other. So, infinite sets can stretch even in size. An even more drastic example is to project a vertical interval onto the whole half line:



The crucial argument that will go against this amorphous sameness of all infinites starts with regarding a point sequence P_1 , P_2 , . . . in the [0 , 1] interval. We claim that such sequence can not contain all points there. This doesn't sound surprising at all, in fact our simplest visualization of a sequence approaching a point makes this claim seem almost obvious. Indeed, the in-between points are left out. Of course we could add any of these to the sequence for example to its start. In other words we can make the sequence not be monotone rather jump back and forth. It still feels absurd that we could add all points of [0 , 1] to a sequence. It's just too much to swallow up.

A very convincing and visual explanation for this could go exactly with using the decimals again: The [0 , 1] points are the decimals without whole parts so the decimal value after the decimal point is exactly the ten equal segments there. The first point of our sequence P_1 is in one of them. This is not quite true because it can be the separating point of two segments. And indeed for example: .7 9 9 9 9 . . . = .8 so this point is in two segments, the 7 and the 8 both. Anyway, P_1 is definitely covered by maximum two .1 length segment, that is by a .2 length. Next we can cover P_2 with a finer segment determined not only by its first digit but the second too. Again, ambiguity means that to be sure we need two .0 1 long intervals. So, .2 2 length can definitely cover both P_1 and P_2. Continuing the same way a .2 2 2 . . . length will cover the whole sequence. So, the full 1 length of [0 , 1] could not be covered.

The real shocker is that we could have started to cover P_1 with more precise segments, that is accept its beginning decimal form given up to any finite length, say the first ten digits. Then again we can cover P_2 with arbitrary small segments. So the total cover can be arbitrary small too.

Usually this argument is made by avoiding the decimals and rather start with an arbitrary small ε length that we cut into pieces as: $\frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \ . \ . \ . \ = \varepsilon$.

Then we use $\frac{\varepsilon}{2}$ to cover P_1, $\frac{\varepsilon}{4}$ to cover P_2 and so on.

Now that we proved precisely that [0 , 1] can not be a sequence of points, it is obvious that the whole half-line can not be even more. But we sequenced all the fractions and so they can not be all the points either. This means that some point that is infinite decimal must be a non fraction.

You might say, big deal, we already knew that from simply that the fractions are all periodical.

In fact, this periodicalness only regards the end of the decimal form so it is showing non fractions within arbitrary small beginning interval. To defend the above argument we can show that it also implies this same. Indeed, the covering of the sequence was arbitrary small, so not only [0 , 1] but any fix interval should be unsequencable.

These two facts that the fractions are easy to sequence all and any sequence is easy to cover with arbitrary small total length are creating a visual contradiction. How can these same arbitrary small covered fractions be everywhere? The weirdest is that both visions are effective too. We can trace all the fractions. How this simple paradox remained unexplored before Cantor, is a mystery!

But the real surprise is that this wasn't Cantor's crucial idea either. And maybe the common explanation is simply Formalism, the obsession with exact proofs. And indeed, in spite of the paradox being concrete, its use for the actual proof of the irrationals lacks the exactness.

The two separate easy and concrete facts that the sequences are arbitrary small coverable and that the fractions are sequences, only imply the irrationals with a third assumption namely that an interval say [0 , 1] is not arbitrary small coverable. In fact, it shouldn't be coverable by smaller length than itself. Though this seems obvious, it should be itself proved jus as convincingly as the other two parts. This came only by Borel who refined the point set arguments of Cantor.
We'll return to his heuristic result much later to see its depth fully.
But can we concretize a missing point from a sequence in [0 , 1], without Borel's direction?
It's quite easy! Remember the two segments that definitely covers a point. All we have to do is pick some segments outside these two. We have eight such, so choosing one is always possible.
But there is a crucial trick required! The outside one of the P_1 covers is free but then the outside ones of the P_2 covers must be chosen only inside that we chose for avoiding P_1. This is not a problem but reduces the choices. If for example P_1 was covered by the 7 , 8 segments and so we chose say 3 as outside segment then we have to chose the next outside of P_2 within this 3.
So if P_2 is . 5 4 . . then P_2 is covered by the two segments . 5 4 and . 5 5 so all other . 0 1 length segments will avoid P_2. But we go into . 3 and avoid . 3 4 and . 3 5, so chose something else say . 3 8. Thus we get a more and more refined decimal that will definitely avoid P_1 , P_2 , . . . all.
The reason we always used inside sub segments is that this way we got a continuing decimal, but geometrically it is also justified as the point inside the nested intervals.
This wider picture was explored by Cantor himself but even this contained some traps though not as complicated ones as Borel's investigation of the coverings. Namely, we can not claim that any intervals nested in each other always hide a common point. If for example we regard an interval without its right end and then regard always the halves namely the second right halves, then these will narrow down to the missing end point. So there will be no common point of all the infinite many, in spite of they all having points and being nested into each other. Of course, we can easily avoid this counter example, by regarding our narrowing segments closed from both ends.
But these are complications again. So, is there a way to go even clearer? Yes there is!
We simply have to go out of the line and regard the P_1 , P_2 , . . . point sequence as a listing of infinite many decimals without whole parts and listed under each other:

1.) = . 2 6 0 3 7 0 . . .

2.) = . 3 7 4 2 4 3 . . .

3.) = . 9 5 9 3 8 2 . . .

.

.

Now we show that some points that is a decimals must be missing from our list.
We create a decimal that can not be the first above because it will not have 2 as first digit, also it can not be the second because its second digit will not be 7, then again it can't be the third because it will have a third digit not 0 and so on. So, we simply avoid all the digits in the diagonal. In short, our created number is an anti diagonal. This gives lots of choices by simply choosing anything but the diagonal digits. If we want to be specific we can choose the smallest other digits from the possible ten digits or the biggest or usually just add 1 to the diagonal one but for 9 it should mean 0. For above this would start as: . 3 8 0 . . .
This can't be on our list. They usually call this a diagonal argument though it is anti diagonal.
The diagonal number . 2 7 9 . . . could very well be somewhere on our list.

We already showed that the fractions are sequencable. The small fractions lying in $[0 , 1]$ have smaller numerator than denominator and so are even easier to sequence as:

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

Some fraction values repeat like $\frac{2}{4}$ has the same value as $\frac{1}{2}$. If we wish we can delete these but it's not necessary. The next step still works if have repetitions. Namely, we replace each fraction with its decimal form and write these under each other. All finite or periodical infinite decimals of $[0 , 1]$ will be there on this list. The finite ones can be continued with infinite many 0-s to make all infinite. Then the above explained anti diagonal, missing decimal thus can not be a fraction.
We might say again that this was the stupidest argument. Going through all the trouble to show that there is decimal that is not periodic. But we have to realize that we didn't rely on periodicalness at all. This was a side issue. The actual listing of the decimals was made by the above sequence with increasing denominators. Of course, we could have listed the periodical decimals by their two parameters, the s starting segment and the r repeating one too. These can be regarded as natural numbers and by increasingly we can list in groups all of them. For example the 100 group would contain all decimals with both s and r being under 100 which is 10000 possibilities. But the point is that we didn't need knowledge of the periodicalness or any features of the digits.
This becomes clear if we apply our method for the algebraic numbers:
Every equation can have only finite many roots, so to sequence all the algebraic numbers is enough to sequence all whole equations. Gödel's method actually was mentioned only for tuples of naturals but works for wholes too and there are even simpler ways to list all these.
So, all algebraic numbers of $[0 , 1]$ can be listed under each other again and the anti diagonal example gives a non algebraic number there instantly.
To emphasize even better the grand universality of this method, we should regard the anti diagonal example in a whole new sense. Not as an infinity of examples that always lies outside any sequencable set of decimals or points of $[0 , 1]$, rather as a counter example to prove that the full set of decimals that is $[0 , 1]$ is a non sequencable set.
Indeed, the full set of decimals can not be sequencable because then our argument becomes an indirect argument. Namely, assuming sequencability, the anti diagonal is a contradiction.
This became the big picture not just by regarding the decimals or points of an interval as a non sequencable set, but later in the generalizations of the whole anti diagonality.
Back to an other application beyond even the algebraic numbers:
If we accept that the possible rule systems are sequencable too, then the ruled decimals are too.
So there have to be decimals not completely ruled.

Infinities Versus Measure

Now we can be like Galileo and wonder away about a strange fact we mentioned already before and came back again above:
Ruled decimals mean decimals determined by some rules. But "unruled" doesn't necessarily mean random. If we rule every second or every thousandth or even just any infinite many digits, then the sequence is obviously not determined. In fact, it remains undetermined for all other digits. But neither can it be regarded as random because infinite many digits are predetermined. We can go oppositely, let only a narrow but still infinite many places be free and determine all the rest by some rules. The freedom or ruling of infinite many digits are leading to conflicting visions:
First of all, no matter how narrow subset is left open. It still means an infinite sequence of digit choices. So the same anti diagonal argument shows that we must have a same unsequencable set.
But now comes the real surprise about the quite opposite, arbitrary narrow predetermined digits.
In fact we go to the extreme! We won't say what digits must be in these places, we merely exclude one at each. The simplest is to exclude the same fix digit at all, say the digit 7.
To prepare the shocking consequence of this, is best by asking someone the simpler question: "What percentage of numbers don't contain the digit 7 up to a million?". Most would say ninety or eighty but the truth is less than fifty. How can the exclusion of one digit exclude more than half of the numbers? Well, up to 10 we have indeed ninety percent 7-less numbers, namely all other digits and 10. But then up to 100 we have at once only ninety percent if we exclude the seventies as a 10 block plus in the remaining 10 blocks we can apply our previous logic that keeps also ninety percent of those. So $.9 \bullet .9 = .81$ that is 81 percent remains.
Similarly, up to 1000 we will have $.9 \bullet .9 \bullet .9 = .729$. So, after six applications, that is up to a million we indeed will only have less than $.5$ that is 50 percent.
But with the infinite many 7 exclusions, the argument is even simpler and yet stronger.
Imagine the $[0 , 1]$ interval again. The possible digits are repeated ten sections within each other. The exclusion can be visualized as blackening out the points that represent those points that contain 7. No matter what the first place is where we exclude the first 7, it will mean blackening out one tenth of the line. If it's the first digit then a full one tenth interval but if it is say the second digit then simply one tenth of all the one tenth little intervals, which together is still one tenth in total. Then the second exclusion is again immaterial where happens, it will exclude one tenth of the remaining white sections. Or to look it negatively, the white areas will always be reduced to their ninety percent. So this will approach zero! Only a so called nil set remains.
So, the amazing fact is that having infinite many free digit choices in decimals will always create same infinity as all decimals that is the $[0 , 1]$ continuum, but locking in, or even just allowing restricted freedom of choices in infinite many digits creates an arbitrary small coverable nil set.
This of course at once means that we can combine the two, that is allow infinite many free choices with also infinite many restricted or locked in digits, and thus we create a continuum nil set.
These strange sets became called Cantor sets but they hide a deeper meaning:

The totally ruled decimals are clearly as many as many rules are. If we accept these by rule systems or machines which are determined by finite sets of even simpler fix rules, then these all are a sequencable set.
But the interesting set is not these totally ruled decimals, rather the strange ones that are not totally ruled only partially, and the complement set, the random ones that are totally unruled.
Since allowing infinite many free digits means continuum many decimals, thus obeying some fix strangenesses can already allow continuum many decimals. In the random decimals the variability of the digits should be more than in the strange ones but strangely we can not guarantee infinite many free digits. So this is a dead end to establish that the randoms are continuum or even exist.
If we knew that the decimals ruled by any fix rule, have infinite many restricted digits then we could at once conclude that they are a nil set. Then using the assumption that we have only a sequence of possible rule systems, all the ruled decimals together were a nil set too.
Then the complement, the randoms were a 1 sized set.

Of course it's not true that a rule defines infinite many restrictions in the digits, so seemingly this is a dead end too. But a concept of restriction can be generalized to beginnings instead of actual digits and so the nil size still follows from some rules.
Thus the 1 size of the randoms also follows.
To define randomness means to define the rules that are to be accepted as strangenesses.
The common feature is always that these strangenesses are nil sets.
So a more simplistic approach merely accepts the nil size as an ad hoc condition.

Behind all this lies the disharmony between the biggest infinity, the continuum on the line and the smallness of the nil sets.
These non exact arguments obviously relate to the already mentioned direction of the coverings, explored by Borel. He went further to define more accurate sizes of sets between 0 and 1 and so initiated the concept of measures. This relates again to probabilities.
The basic failure of measures is that they can not be complete. Some sets are non measurable.
As we'll see this is not caused by the above disharmony.
Also, the infinites have their own failures explained next:

The Continuum Hypothesis And Russell Paradox

This disharmony of infinites and point sets is especially important from an other angle.
Indeed, Cantor raised the most logical question after he saw that the points of even the smallest interval are more than the naturals or fractions or algebraic or ruled numbers and so on.
Namely, is there some new infinite between this simple sequencable infinity and the continuum, the points of space. He thought there shouldn't be and called this the Continuum Hypothesis. Now, we might think that the examination of point sets and maybe measures hides the answer.
But things turned out to be different.
Paul Cohen proved that Set Theory can not decide this problem.
The hidden past behind this new result is the period between the naïve original Set Theory and Gödel's first grand application of it for the Completeness Theorem.
This was the axiomatization period that contained two big surprises.
One I already mentioned, the Axiom Of Choice implying the Wellordering which was used in the Completeness Theorem too. The other one is actually relating to the Incompleteness Theorem.
Indeed, it was the first reoccurrence of Cantor's anti diagonality in a language form.
Incompleteness became the second. Just as here the self reference is regarded as use of the "liar paradox", the original logical appearance was also labeled as paradox. Namely, the Russell Paradox. Though it has absolutely no conflict with our plausibilities.
If we regard sets as collection of objects then $\{ x \ ; \ . \ . \ . \}$ should be allowed to collect all those x sets that satisfy the property described by the dots. If we try to collect all sets that is form the universe of Set Theory then we could use $x = x$ as the collecting property because this should be true for everything. So $U = \{ x \ ; \ x = x \}$. The elements of such collections are those sets that satisfy the property of collection. Now $U = U$ too, so this implies that $U \in U$.
This seems pretty strange that something is element of itself but not contradictory yet.
Lets try to collect on the other hand all those sets that are not such controversial ones that is are not elements of themselves. The total N set of these normal sets is then $N = \{ x \ ; \ x \notin x \}$.
Now, is N itself normal or self elemental? Well, if it is normal so not element of itself then
$N \notin N$ so it satisfies its collecting property and thus should be element of itself.
If it's not normal that is self elemental then $N \in N$ so it doesn't satisfy the collection property and so it shouldn't be element of itself. We get contradiction either way.

The Fourth Original Approach And Finishing The Third

The above Russell Paradox was long buried and forgotten and most importantly avoided by trickier collection axioms when the new diagonality appeared by Gödel.
At first it was the language meaning again, the self reference or paradox formulation that seemed as more important. But the anti diagonal meaning was also clear at once too.
What we above described as object assignments to derivation systems, was originally the assignment of numbers to the expressions of a language. The famous Gödel numbering. And Gödel truly had to go through and create such numbering. Nowadays we just imagine it being done. Because the point is that then the language can talk about its own derivabilities. Some jump to final statements that say "I am not derivable" but this is a cheap over simplification. The much better vision is to see that there is a $U(x, y, z)$ formula that can express the derivation of $P(\ )$ properties at some n cases. Namely, x means the number assigned to $P(\ )$, y the number we represent by some names and use in P as case and finally z the number assigned to a derivation.
If $[e] = E$ means in general that e is the assigned number to the E expression then $[p] = P(\ )$, $[n] = N$ and $[d] = D$. So, $U(x, y, z)$ at $x = p$, $y = n$, $z = d$ that is $U(p, n, d)$ "means" that $P(N)$ is derived by D. The most obvious problem is that a D derivation is not a purely language concept. It depends on what axioms are used. Accepting this dependence of the assignments for derivations from the axioms, a second problem becomes what the word "means" above for $U(p, n, d)$ should itself mean. The two options are that it should be derivable that is being theorem of the axiom system or rather being true among the natural numbers.
Using a simpler $U(x, y)$ that expresses the existence of a $[d] = D$ derivation and simply using $N = n$, that is the naturals as names for themselves, the two possible meanings are:

$U(p, n)$ theorem $\leftrightarrow$ $[p](n)$ theorem or $U(p, n)$ true $\leftrightarrow$ $[p](n)$ theorem

Both will instantly prove that there has to be undecidable statement in the system. Namely, in both options the $\neg U(x, x) = Q(x)$ property will provide the undecidable statement, at its own code as case. So if $\neg U(x, x) = Q(x) = [q](x)$ then $[q](q) = \neg U(q, q)$ is undecidable. Indeed:
In the first option: $U(q, q)$ theorem $\leftrightarrow$ $[q](q)$ theorem $\leftrightarrow$ $\neg U(q, q)$ theorem.
If the two ends were true it would mean a contradiction in the system. So if the system is consistent then both ends are false which means that neither $U(q, q)$ nor $\neg U(q, q)$ is theorem. That is, they are undecidable.
In the second option: $U(q, q)$ true $\leftrightarrow$ $[q](q)$ theorem $\leftrightarrow$ $\neg U(q, q)$ theorem.
Now consistency is not enough, only the assumption that all theorems are true. Indeed, then the two ends can not be true and so their falsity means that $\neg U(q, q)$ is a truth that is not theorem. But also, $U(q, q)$ is not true and so $U(q, q)$ can not be theorem either.
The anti diagonal feature is evident.

These built in universalities are the fourth but actually original approach of Gödel.
The existence of such U is hard to prove and has to be shown for every axiom system separately.
The earlier unfinished third approach is the total opposite. It works for all consistent extensions of the raw system. The representability of the derivable sets in the raw system means that the totally derivable ones are perfectly representable. This inherits to the consistent extensions.
Then in all such consistent extensions we can argue to show undecidable statements.
We would expect to use again the perfect representability of the totally derivable sets, but we are wrong! The perfection was only needed for the inheritance. The argument in the extended system only needs that all totally derivable sets are representable. This already can imply undecidable statements. But not directly!
We only show first that the set of theorems and non theorems can not be both derivable. Indeed:
Suppose a T system could derive the theorems and an N the non theorems.
Lets put all the $P(x)$ properties in order by our text alphabetical method and let $|P|$ denote the position in this order as a unique number assigned to P.

By T and N we can create T' and N' systems that derive exactly those |P| numbers where P (|P|) are theorems or not theorems respectively.
Indeed, they both can recognize among their derived statements the P (n) formed ones and among these also when an n is identical with |P|.
Let S be the set derived by T' and so ¬ S the one derived by N'.
These are totally derivable sets and so should be both represented by some properties.
Suppose Q represents ¬ S. If Q (|Q|) is a theorem then |Q| should be in S because S was the cases of the P (|P|) theorems. But |Q| should be in ¬ S because Q represents it.
Similarly, if Q (|Q|) is not a theorem then |Q| should be in both ¬ S and S .
So the full representability of all totally derivable sets and the assumption of the theorems and non theorems both being derivable sets, led to contradiction.
So, if we know that there is full representability, then the theorems or the non theorems must be a non derivable set. This still doesn't imply incompleteness directly. For example, we could separate all the statements by their truths and accept all true ones as axioms. It's a complete system but neither the theorems nor the non theorems are a derivable set.
If one is derivable and the system is complete then the other must be too because the formal negation makes that derivable too. But we just proved that they can not be both derivable, so if one is derivable then the system can not be complete. Usually we have derivable set of axioms and so the theorems are the derivable and the non theorems must be non derivable.
The clean beauty of this third approach shows how much easier things became by simply seeing what's going on. The same anti diagonal argument that had to go through a messy particular proof at Gödel, now gave much deeper explanations. The basic point is what I emphasized already a few times: Instead of chasing a "concrete" undecidable statement, the non derivability of the non theorems is proved. This can inherit from our raw system to extensions with some assumptions.
Quite trivially to all extensions that are true among the naturals, this was our second approach.
Or if the raw system is extended with $<$ then the perfect representability of the totally derivable sets will inherit to all consistent extensions and we don't even need the external assumption of an underivable complement. We create one in or rather above the system as the set of theorems and non theorems. And we don't have to use this fact inside, it proves the incompleteness at once.
This was our just finished third approach. The fourth approach is the original road avoiding the concept of derivabilities outside the axiom system.
A counter argument against external derivabilities, that is for the fourth approach could be that if we want to be precise we need a concrete framework of derivation systems or other effectivities for the outside and that needs the extra proofs. Plus all this outsideness makes things a bit fishy.
The first part is true, we need extra details but these are not fishy at all. The systems of our chosen framework must be representable in our axiom system and our derivations in the axiom system must be representable in that framework. So the framework could be eliminated and argue solely in the axiom system. But this would be much messier and couldn't apply to other axiom systems.
All this shows the much deeper common thread no matter which approach we choose, namely the reoccurrence of anti diagonality. I will prove this point even further in the next two sections.
So, self reference which is only an explanation in axiom systems, is a false cause in my opinion.
But as I said earlier, I might be wrong if the field of effectivity can prove a hidden negation behind all effectivities. Anti diagonalization is already a visible tendency there.
A first proof of this starts with going into listings in general as I promised.
Above the assignment of numbers to the P (x) properties was very simple, almost natural by the text alphabetical order. To carry out a full assignment for all expressions and derivations is much harder. The easy assignment above worked because we had the strong indirect assumption of the two complementing systems T' and N'.

Weird Listings And Anti Diagonal Like Arguments, Turing's Anti Diagonality

We might think that above already the T' system that was able to derive all totally derivable sets shouldn't exist and so should be enough to get a contradiction. But it's not so.
This feeling is based on exactly the simple assignment. But we couldn't use that simplicity as tool, so we assumed the assignment as practically arbitrary and instead we used the heavy gun that this assignment gives the complementing sets in tandem. Listing is a better vision than assignment.
Listing is the concept between mere sequencing of lines under each other that are themselves mere sequenced sets, and lines created by programs through a universal machine.
The first is Cantor's original vision, the second is too hard to tell exactly because being program is a complicated fact. So we need the in-between concept of listings.
Here the lines are created again by a machine but not by programs merely a parameter.
If the objects are natural numbers, we simply have an L (p , n) machine that has a split input and thus can create a listing. The p program like parameter will tell the lines and with a fix p value the n tells the elements in the p-th line. How program like is p, depends on how much we assume about L. We mentioned the three level of universality. That is, L being universal, p being universal in the sense that we get all derivable sets by using some fix values as p, that is as lines, and finally the crucial third level is if p can imitate other splittings of the input.
The tuple input and number valued functions that Kleene used as base of effectivity, were defined in ways that obey this third level of universality. That's why the program like results all follow from his famous theorems that actually hide this third crucial universality. But we don't always need this full program like assumption of p.
The goal is to create contradictory listings or create weird ones that we would think is impossible.
This second is always much harder than the first that simply boils down to some anti diagonal like argument. An example of the second one is what we revealed above, that the T' listing of all totally derivable sets is not contradictory in itself. So, there is a weird listing of all totally derivable sets. Obviously it is more tricky than the simple text alphabetical T'.
Instead of showing this, we go towards the easy and more enlightening first examples, that is anti diagonal like arguments to prove that some listings are impossible. We give two examples. These show that finite sets have a crucial role among the effective sets. Excluding them, that is regarding only infinite effective sets we can easily create anti diagonal impossibilities:

1.) We can not list all and only the infinite derivable sets.

2.) We can not list all and only the infinite totally derivable sets.

3.) We can not list some infinite derivable sets that contain all totally derivable ones.

Seven important points:

The "only" criterias were added merely to emphasize this fact. Just like derivabilities always mean exactly deriving a set, listings also always mean exactly listing a class of sets.

The third claim obviously implies the second, so enough to prove the first and third.

The first is surprising because we <u>can</u> list all derivable sets, namely by their programs using a universal machine.

The second is surprising because of the mentioned fact that there is weird listing of all totally derivable sets.

This second result of course raises the question whether instead of restricting the contents to infinity, restricting the listings could also show impossibility. Namely, a program like listing of all totally derivable sets should be impossible. It is, but much harder to prove. Even the listing of all actual programs for them doesn't imply trivial anti diagonality, because we don't know which programs are complementing. Listing the programs in complementing pairs of course gives anti diagonality at once, exactly as above the T' , N' pairs did.

The lines will be regarded as recognizing numbers from the sequence: 1 , 2 , 3 , . . .
Namely, as lighting up and staying lit up. This is very visual but don't forget that the appearance of these light ups are not increasing by the numbers. A small number can light up arbitrary late.
Luckily, since we know that we'll have infinite many lit up numbers, we can always wait.
This is the whole essence of the two proofs.
In both proofs we give our own recognizing visualized as circling some naturals. These will be done in increasing order so it's not merely a recognizing but at once gives a recognizing of all the non circled ones too. Indeed, once we circle a number, we can recognize the new complement members as the numbers since the last circled one. So, we create a totally derivable set.
For the third claim this has to be so but for the first it's an extra topping on the cake.
1.)
We watch the first line and whatever number is first recognized that is lights up, we circle, the next one. This will be our smallest, so the first line at once can not be our line because it had the previous number in it. Then we wait until in the second line a number lights up that is bigger than our first one was. We will again circle the next number, so definitely skip through the one lit up and so our set can not be the second line either. And so on, we collect numbers in increasing order and all lines are made sure not to be identical with our.
3.)
Now we wait for the first two numbers to light up in the first line.
We put the smaller into an S basket while the bigger into a B.
Then in the second line we wait for two numbers to light up that are both bigger than either of the two above found numbers in S and B. We again put the smaller in S, the bigger in B.
And so on, we always wait for two new values bigger than any numbers in S or B and put them in the baskets properly.
The total of both baskets are clearly totally derivable sets because they are increasingly generated.
They are disjoint that is contain no common numbers. But our lines all contain numbers from both and so none of them can be identical with S or B. So S and B are missing from the totally derivable sets among the lines and thus those appearing couldn't be all possible ones.

Finally we have to emphasize that the first who extracted the pure anti diagonality from the foggy self reference and paradox vision of Gödel, was also Turing. That's why I waited till now to finish his whole argument. We showed that there is universal machine but we didn't show the final step, the underivable complement arising from it. Of course by now probably it seems obvious as a diagonal use of it. But I will show a surprise. Namely, we can show a much simpler feature of this diagonal universal machine that easily implies the underivable complement:
A t input is a shared input of the M , N machines if either they both stop or both run from t.
The M-stop or M-run sets are those inputs from which M will stop or run.
Two M , N machines are complementing if the stop set of one is the run set of the other.
Obviously, if M , N share at least one input then they can not be complementing.
If an S machine shares at least one input with every M machine then S has no complementing machine.
Now all we need is such S machine that shares input with every machine.
The U' (p , t) universal machine using the special symbols for both p and t can be used with same members. This U' (t , t) then simulates machines running from their own program as input.
Thus, this U' (t , t) = S (t) is sharing input with every M. Indeed, if M's program is t and it stops from t then S does too but if it runs then S does too.

The Fourth Pillar

The first three pillars of New Math are Set Theory, Logic, and Effectivity!
Set Theory was Cantor’s ingenious recognition and became the foundation because all math can be formulated as truths about structures that is sets with inner sets. That was not Cantor’s initial motivation rather the discovery that sets have definable sizes. These infinities of sets are not that crucial at first sight in the structures but looking deeper they always appear.
The structures are usually models that is made for axiom systems. Logic tries to derive the truths in the models. The success of our Logic was the crucial claim that not only the fairly obvious fact stands that truths obeys logic but in reverse too logic obeys truths. Consequences of some axioms that are true in all models are derivable by Logic. That was Gödel’s Completeness Theorem.
Our Logic is complete.
His Incompleteness Theorem claimed the incompleteness of most axiom systems, so the word completeness had a totally new meaning here, namely decidability of all statements.
The imprecise “most” means two things. One that the axioms must be not just given as a set but rather as a generable set of statements, and the second that the derivable statements must be at least as complex as a simple number theory. This second condition was over emphasized first and thus the language role too and regarding the claimed undecidable statement as a kind of paradoxical self statement. Turing was the first who saw the first condition as much more important. The generability of the axioms means that continuing with logic the whole set of derivable statements that is the theorems is merely a derivable set in a general sense as set of objects. The non theorems the left over statements or the complement set is then can be seen as a much more complex set that is not derivable not just in the logical sense but again abstractly as a set by any methods. This of course implies the incompleteness at once because if every statement is decidable then the normal derivations of the theorems and then an added negation would give at once a derivability of all non theorems. So the non derivability of the non theorems as set by any system can only happen if the theorems already don’t pick one from every pair of $A, \neg A$ statements. Strangely, the self reference still remains in this generative or mechanical or effectivity view, because the proof that the non theorems is a non derivable set relies on that the theorems are complex enough that their complement is even more complex. Apart from this, the truth is that most generated sets are complex enough to have non generable complement but to prove this for a particular case requires pretty strong assumptions.
These non effective complements of effective sets or non effective sets in general became an interest on its own and gradually lead to the fourth pillar Randomness.
The obvious sign of randomness being something important are those little remarks we made earlier about decimals. Usually we regard binary $s = 00111010001000100\ldots$ sequences but our earlier decimal ones were just as good.
Being totally determined by effective methods obviously means that the sequence is not random.
But we have firm intuitions about a much wider class of sequences that are not determined and yet not quite free that is random either. So while randomness is a vague feeling of this free or totally undetermined sequence, we have very concrete intuitions about the opposites, when a sequence is not free or random, rather manipulated or artificially influenced. More exactly, for every concrete claim we have our a priori intuition that tells whether the claim is a “natural” expectable property that stands for all random sequence, or it is something “unnatural” that shouldn’t be true for a random sequence. An other word for this unnaturalness could be “strangeness”.
I will distinguish these two because it reveals an important point.
The most obvious widening from effectivity is having a pattern within. So, not all but only an infinite subset is predetermined. Then the sequence is not random any more. We also saw that even just infinite many digit exclusions in base ten mean drastic reductions in chances.
If we regard these sequences as gambling outcomes then the non randomness means possible wins. But this betting scenario is not necessary. Without it we have very good intuitions directly what claims about a sequence should constitute a strangeness and what are merely naturalnesses, that is true for all random sequences.

That's how I turned to Randomness after Set Theory in high school back fifty years ago.
I spoke not a word English and knew nothing about Turing or effectivity.
Yet I practically discovered the whole field on my own primitive way. Now this primitiveness will help you to get a fast view in this section before we go to more systematic details:

The most obvious way to reveal to someone that randomness is a weird feature, is to ask him to tell some things about an imagined s = 00111010001000100 . . . random sequence.
The first possible claims will be about the beginnings and then if we try to push the person to say something about the long run he will indeed come up with some natural claims that are true for all random sequences. Like having infinite many zeros and ones.
Then we can draw the attention to the amazing fact that while the finite or beginning claims were all excluding some differently starting random sequences, these infinite ones were not excluding any, so they are true for all of them. Now, to mix up some finite and some infinite claim is easy! Simplest is to connect them by "and" or "or".
The negation of the natural claims that are true for all random sequences will be "unnatural", true for none. Having ten long all 0 segments infinitely is natural while having only finite many of these would mean that after a point we know that if we see nine 0-s then a 1 must come. Of course we still wouldn't know when this pattern starts but it would be a good strategy to bet on these 1-s because after a while we would win for ever.
Now if we envision instead of a random s sequence an intentionally non random one then first it seems that the same duality will come in. That is obvious finite properties about the beginnings and bigger claims about the infinite behaviors. But looking closer we'll see that now these infinite properties relate quite differently to the finite ones. At the random ones they were very alien to each other and we could only artificially combine them with "and" or "or".
Here a claim like having a pattern is naturally contains already the finite sections.
To have hundred blocks of 0-s and 1-s repeating alternatively, not only is an infinite tendency but tells that the sequence starts as 00000 . . . .
Now it's very easy to turn this pattern into a purely infinite pattern by adding the condition that it stands "from a point".
So this added condition turns any strangeness into an unnaturalness.
The "strangenesses" are the widest possible patterns or trends or irregularities and these are involving the beginnings. The "from a point" condition can take out artificially this beginning determination and create an infinite strangeness. This then is actually an unnaturalness. Its negative is a naturalness. Indeed, not to have fix blocks of 0-s and 1-s from a point is a naturalness. It is true for all random sequences and it doesn't involve restriction on the beginning.
But we feel that these artificial naturalnesses are twisted not "natural" so to speak.

First we will exactify these and then come back to this same basic result that strangenesses should be the start not naturalnesses or unnaturalnesses.
An S set of sequences is cuttable if for every $s \in S$ and for every b beginning of s, the $s - b$ sequence remaining after cutting off b from s, is also $\in S$.
An S set is alterable if for every $s \in S$ and for every b beginning of s, altering b to a b', that is $s - b + b' \in S$ too.
If an S set is alterable then its complement $\neg S$ is alterable too. Indeed, if an alteration $-b + b'$ would lead out of $\neg S$ into S, then its $-b' + b$ alteration would lead back from S into $\neg S$.
If S and $\neg S$ are both cuttable then already they are both alterable too.
Indeed, a $-b + b'$ alteration from S can not get into $\neg S$ because $s - b$ is still in S and $s - b + b' - b'$ is the same but it is also in the same set as $s - b + b'$.
An S set is b-total if all continuations of b are in S. An S is total if it is b-total for some b.
An S is continuable if any b has some continuation in S.
Clearly: An S is continuable if and only if $\neg S$ is not total, that is not b-total for some b.
If S is alterable then trivially it is continuable, namely every $s \in S$ in fact every $s - b$ can be a continuation of any b'. So, also an alterable S can not be total either.

The symmetrically alterable pairs of S , $\neg S$ and the non symmetrical consequences of continuable S and non total $\neg S$ are best represented by the theoretical and practical approaches to randomness.

Theoretically, the best to start with the Law Of Big Numbers by which any b segment can be the outcome of random trials. We can go in order of the segment and if we fail we simply restart but it's much more visual to do the whole segment simultaneously. This way the order is irrelevant and we avoid the usual trap of believing that the past could influence the next outcome.
For example trying hundred coin tosses simultaneously, we'll see that among the outcomes where the first ten are all heads the eleventh being head or tail is the same amount.
For us now the simple point was that any predetermined hundred outcomes can come about.
Of course, if they can come about simultaneously then they can come about after each other too.
Then this return to the consecutive trials can lead to our first plausibility about sequences if we simply continue the trials: Every beginning can be the beginning of a random sequence.
A second intuition is that quite generally too, a random sequence is random from any point.
Any beginning can be cut off. The random sequences are cuttable.
This seems simpler than the previous but actually it claims randomness about the remaining infinite section so it is a more complex claim.
The reverse of it, that we can place an arbitrary beginning before a random sequence is much harder to see because once a sequence is visualized as infinite time sequence, we can not go back to the past to add new trials, especially repeatedly tried ones to get the wanted beginning.
What really helps to convince us that this is still true, is to imagine the opposite of being random.
This is having some strangeness. Since the beginnings are already cuttable we can assume that the strangeness itself is cuttable. Thus of course we actually have both the random and non random sequences becoming alterable and these strangenesses we assumed are infinite strangenesses.
The step from assuming these strangenesses as purely theoretical, merely having some, to the individually existing strangenesses, is the simple empirical fact that we can easily create such.
Any order in the sequence like prescribed digits or repeats can do. Then the previous assumption of infinite strangenesses can also be achieved by simply adding that these stand "from a point".
This empirical concreteness includes the fact that these concrete strangenesses will still exclude all random sequences.
So just as the R set of all random sequences and its complement the $\neg R$ set of sequences that possess any infinite strangeness were obviously two alterable sets, now we assume that a concrete infinite S strangeness and its complement $\neg S$ is again an alterable pair.
S excludes all the random ones so they are all in $\neg S$, that is $R \subseteq \neg S$.
Now comes the final and most crucial assumption that the R set of all random sequences is a special one in respect of any possible alterable collection pairs. Namely:
We can not create any such pair that would split R. This always remains fully outside one of the alterable pair and fully inside the other. The one that excludes R is the infinite strangeness and the one that includes the whole R is merely a naturalnesses.
So, these intuitive concepts we started with, would now have totally precise definition as members of the alterable collection pairs. The only question is which is which. But one single random sequence in one of them tells that it is the naturalness and the other is the strangeness.
A fundamental failure of this vision comes about if we regard the collections of sequences as theoretical sets. Indeed, imagining an s random sequence, we can add to s all the beginning altered variants and then this S set is clearly alterable. Yet it can not contain all random sequences but contains some. So S splits R.
Strangely, this failure doesn't bring any similar problems if we use explicit collections by formulas that is by properties. The reason is simply that such explicit properties can not use a random s as start. This suggests to go this way that is choose a smart theory and define the alterable collections there. There were such approaches to randomness but they died out because a much more heuristic idea came in.
The formal start could be to regard the concreteness that avoids the previous problem not as explicitness rather as effectivity. To collect effectively infinite sequences of course is problematic

because a machine directly handles only finite objects. The full set of these objects is still infinite so there is no absolute contradiction but the set of sequences then must be determined by set of finite objects that is segments as best. One idea could be to use these segments after each other to create the sequences and this is nice because then they all are independent from each other. But we would have to restrict what kind of lengths follow and so on. The other idea is to regard the segments all as beginnings of the sequences. The longer and longer ones can define a sequence but the continuability then is a hidden factor. The amazing truth is that the final method that triumphed is using this beginning vision but quite barbarically, half of the probabilistic arguments behind it assume the independent continuation vision.

The fact that we build the sequences from their beginnings, dictates that we should relax our search from effective infinite strangenesses to merely strangenesses. Indeed, to say that something is true from a point would mean that all possible beginnings must be added as alternatives. It's much simpler just to concentrate on mechanical strangenesses that might depend on beginnings.

The real beauty is that the machines themselves take care of the rest.

For example, a simple strangeness is to have hundred blocks of alternating 0-s and 1-s.

A machine can easily produce this too. Now, altering a beginning ruins the whole property but of course it remains true "from a point". But instead of saying this, a machine can create the basic sequence and then alter arbitrary long beginning from its memory. So, the altered sequences remain machine determined. They simply become a new strangeness by an other machine.

In the end the random sequences are only those that fail all strangenesses, so we get them.

Giving up the alterability means that some b continuations will be excluded from the B collection of possible beginnings. Then from the continuable ones we might again exclude some continuations. The possible continuations are narrowed step by step and the remaining full paths tell the sequences determined by B. This is beautifully simple and seems too simple to cover all strangenesses. The mechanical predictions of course fit right in and even the so called conditional predictions. Like claiming that after certain beginnings some situations must follow. Indeed, we can turn these into beginning continuation exclusions.

The really doubtful strangenesses are the probabilistic ones. These don't claim that after certain beginnings something must come, rather claim that something is more likely. Knowing such deviations from randomness of course would also mean that we can bet on the unnaturally higher probability outcomes and win in the long run unlimited money. How could a simple selection from beginnings include such deviations in chances?

To see this, we should realize that building an s sequence from our B set can also be said by simply claiming that our s sequence has infinite many beginnings from B.

The negative, that is not being able to create s from B, means the exact negative that is having only finite many beginnings of s in B.

This then implies that, if an s is not obtainable from B then it is obtainable from $\neg B$.

So we only have to see that that the probabilistic claims are indeed infinite or finite occurrences from beginning sets. Then we can choose the infinite occurrence, that is continuability as the strangeness and the non continuability, that is finite occurrence as the non strange that is natural.

This will give us also a very crucial insight into a fundamentally false view about these probabilistic claims. Usually the non strangeness, that is naturalness is regarded as an infinite claim. So, to see that these are actually stoppings is quite paradoxical.

But in fact only a systematic misrepresentation will be destroyed.

The common name of these normalities is the Law Of Large Numbers.

The relative frequencies, that is the ratio of the n successful outcomes and the total N trials, tends to the probability: $\frac{n}{N} \rightarrow p$

For coin flips, this means that both the heads' and tails' ratios tend to $\frac{1}{2}$, so they "equalize" too.

For a dice, the six outcomes' ratio tends to $\frac{1}{6}$.

The fact that we have to divide, already hides the lie, we are not dealing with a plausibility.

But for the coin flips, we can see quite directly that the head tail “equalization” is a major over simplification as follows:
The earlier mentioned Law Of Big Numbers is indeed a plausibility, saying that everything that has a chance, can happen by random trials. So any segment of head tails is also possible. Thus, if we envision fix, say 100 long consecutive windows in a flip sequence, then all possible 100 long segments will appear in the windows. In fact, infinitely often. Including all heads or all tails too. These windows can be arbitrary long, plus could be placed not just consecutively but at any pre-determined places. In fact, we can even place them at positions that are determined by the outcomes up to the window. Indeed, the past outcomes can not influence the following ones and so in these windows, we should still obtain all possible segments infinitely.
So, assuming that both the heads and tails can become more alternatively infinite many times, we can place the windows after the infinite many exact equalizations. The beginnings ending with these full head or tail windows can create arbitrary big excesses of heads or tails.

But back to the final death blow to the Law Of Large Numbers, we have to really look behind what $\frac{n}{N} \rightarrow p$ means. The first simplification is $|\frac{n}{N} - p| \rightarrow 0$. So in fact, an approach to 0 or so called diminishing is claimed. This means that $|\frac{n}{N} - p|$ remains under arbitrary ε value from a point. That is: For every ε there is an M that if $N > M$ then $|\frac{n}{N} - p| < \varepsilon$.
This exactly means that the $|\frac{n}{N} - p| \geq \varepsilon$ property of N is stopping.
Of course, this property has not only the N variable, but ε too. This means that actually we have an infinity of these properties, for every fix ε value and they all have to stop for N.
The Law Of Large Numbers is a normality or naturalness, true for all random sequences.
And it claims the finite occurrence or stopping of those N values that $|\frac{n}{N} - p| \geq \varepsilon$.
Quite on the contrary, the infinite occurrence of such N-s is a strangeness, the failing of the Law Of Large Numbers.
Thus we also showed that our strangeness concept of continuation or infinite occurrence from a B beginning pool is covering these probabilistic strangenesses too.
But now comes a death blow to our own idea!
We can mechanically list all possible beginnings as:
0 , 1 , 00 , 01 , 10 , 11 , 000 , 001 , 010 , 011 , 100 , 101 , 110 , 111 , 0000 , . . .
As you can see, we went in increasing length and within by increasing binary values.
Now, from this total pool, every s sequence is continuable, so all sequences would be strange and no sequence could be random.
We have to restrict B in a manner that avoids these too flexible collections!
Observe, that even if we just allow all continuations of a b beginnings, then obviously we would allow random sequences, because every b can be continued randomly.
Omitting beginnings from the continuations of every b would avoid such fix bundles that trivially contain random sequences, but the solution to close out all of them requires much stronger narrowings. And yet the final solution is amazingly simple:
We have to add up the chance values of all b elements in B and require that this chance total of B is not infinite, only finite. The crucial consequence of this finiteness is that if we use up more and more beginnings from B then not only the remaining ones diminish but actually the total of the remaining ones diminishes too. This then means that if an s sequence uses up infinite many from B then also the total chance of the beginnings longer than an N is diminishing too as N grows.
Now comes an easy fact that the chance of “or” connected events, that is at least one coming true from a few, is always maximum the total sum of the individual chances.
Indeed, if the events are excluding each other then the “at least one” is exactly this sum but if some of them allows some of others then that reduces the chance of at least one coming true.

Using this for the infinite many beginnings longer than an N, the chance of at least one coming true is maximum the total chance of the beginnings after N.
But this total is diminishing and so the chance of at least one occurring after N is also.
Now comes the crucial step!
The claim of infinite occurrence implies such at least one occurrence after every N.
Also, a primary claim that implies another, can not have a larger chance than the implied claim.
Here the infinity is the primary claim and it implies the infinite many existing occurrences after every N. These have diminishing chances as we established above and so the primary infinite occurrence can have only 0 chance.

A 0 chance doesn't always mean impossibility for a random event!
For example, a dart landing on a point of a board has 0 chance and yet always comes through.
But here, with repeated infinite trials, we regard this 0 chance for a sequence as an impossibility.
Thus, indeed for random sequences, such finite total B must stop. Infinite occurrence is a strangeness.
So finally:
An s sequence is random if for every finitely, that is machine generated and finite totaled B beginning pool, s stops in B that is has only finite many beginnings in B.
So again three finiteness means the crucial definition just as at the Taboo Avoidances earlier.

Probabilities

Probabilities or chances are assigned to events. An experiment is a full set of elementary events or cases or outcomes of the experiment. These cases are the possibilities, excluding each other and the fullness means having a 1 total value of the chances of all cases.
The simplest experiment is flipping a coin. The head and tail are the outcomes with half chances.
A dice throwing has six cases with $\frac{1}{6}$ chances. An experiment can have infinite many cases too.
For example, at throwing darts onto a board the cases are the individual hits into points.
These cases have all zero chances and only non elementary events will have positive chances.
The total of chances here means integration and that's why the infinite many zero can add up to 1.

The non elementary events are any sets of cases meaning "or". For example:
To get an even throw with a dice is the event that has the 2 or 4 or 6 cases.
Throwing a bull's eye, that is landing in the inner smallest circle on the dart board is the event containing all those landings as cases.
The chances of these events are the totals or integral of their cases.
The complementer event contains all the outside cases and has $1-p$ chance.
The "and" and "or" combination of events simply means the common part or union of the events as sets of cases.

In simple experiments we have finite many cases with equal chances and so the calculation of the chances boils down to case counting. The number of the desired cases divided with the total case number gives the chance. What makes this not so obvious, is that we sometimes have to create the experiment from simpler experiments.
A repetition or simultaneous versions of a single experiment is the simplest combining of experiments. For example, throwing a dice twice or throwing two dices.
These are independent experiments. They don't influence each other's outcomes and so the "and" cases from the two experiments will have the product chances.
For example, to throw two sixes has $\frac{1}{6} \bullet \frac{1}{6} = \frac{1}{36}$ chance.
This is one of the 36 cases in the combined experiment.
An event in the combined experiment is for example the throws with more than 10 total.
The cases are: $5+6$, $6+5$, $6+6$ that is three, and thus the chance is $\frac{3}{36} = \frac{1}{12}$.
Sometimes the combining from experiments is not simple repetition.
For example, pulling cards from a pack is not mere repetition because after a pull we have less cards to pull from. Thus pulling three cards has not $52 \bullet 52 \bullet 52$ many cases only $52 \bullet 51 \bullet 50$.
Pulling three aces similarly has $4 \bullet 3 \bullet 2 = 24$ cases and so the chance of this is 24 divided by $52 \bullet 51 \bullet 50$. But the counting of the desired cases can have some further tricks too.
For example, pulling at least one ace from three pulls might seem logical by regarding three situations. Namely pulling 1 or 2 or 3. In the first case, the ace can be the first pull which has 4 possibility and the other two pulls 48 many. This is $4 \bullet 48 \bullet 48$ possibilities but similarly the ace can be the second or the third, so we have $3 \bullet 4 \bullet 48 \bullet 48$ possibilities only in this first situation. We have to add the two ace and three ace situations. As we see this is quite complicated.
A much easier way is to calculate the cases of the complement feature which is not pulling any ace. This has simply $48 \bullet 47 \bullet 46$ many cases. And so the desired case number is :
$52 \bullet 51 \bullet 50 - 48 \bullet 47 \bullet 46$ and dividing this with $52 \bullet 51 \bullet 50$ gives the chance.

Independent "or" Formula, Sum Bound Of "or", Bound Chain

Independent experiments have "and" cases with product chances.
For example two dices landing on 6 has $\frac{1}{6} \bullet \frac{1}{6} = \frac{1}{36}$ chance.
The analogy would be that "or" that is at least one landing on 6 has $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$ chance.
This is false and would be absurd by continuing the logic to six dices giving a 1 total, that is certainty of at least one 6 from the six dices. We can throw a million dices and still not get any 6 though this has extremely small chance. So what is the correct formula?
Just like earlier the complements were useful to simplify case countings, they lead us to the solution here too. For example, the correct cases of the at least one 6 landings from two dices are (6 , 1) , (6 , 2) , (6 , 3) , (6 , 4) , (6 , 5) , (6 , 6) , (1 , 6) , (2 , 6) , (3 , 6) , (4 , 6) , (5 , 6) = eleven cases and so the correct chance is $\frac{11}{36}$. But more importantly, the complement has $\frac{25}{36}$ chance.
And this is trivial because not having 6 means having only five possibility in both dices.
Even better expressed, the chance of not having a 6 is $\frac{5}{6}$ and this does multiply to $\frac{25}{36}$.
So, in general if **1 , 2 , . . . , n** are independent events with $p_1 , p_2 , \ldots , p_n$ chances, then the complement chances are $(1-p_1) , (1-p_2) , \ldots , (1-p_n)$ and so no occurrence at all has the chance $(1-p_1)(1-p_2) \ldots (1-p_n)$ and so the at least one occurrence has $1-(1-p_1) \ldots (1-p_n)$.

Our false sum analogy above was not quite meaningless.
It is a bound on the true chance regardless how the events relate to each other:
$p(\mathbf{1} \text{ or } \mathbf{2} \text{ or } \ldots \text{ or } \mathbf{n}) \le p_1 + p_2 + \ldots + p_n$
Indeed, if the events all exclude each other than equality stands but if some can happen together than this reduces the chance of "or".
This of course means that this bound is true for independent events too and so:
$1-(1-p_1) \ldots (1-p_n) \le p_1 + p_2 + \ldots + p_n$
We can prove this purely algebraically too, with induction:
For $n = 1$ the equality stands. Then assuming it is true for $n-1$ and multiplying both sides with $(1-p_n)$ and adding p_n to both sides, we get:

$$(1-p_n) - (1-p_1) \ldots (1-p_n) + p_n \le (p_1 + p_2 + \ldots + p_{n-1})(1-p_n) + p_n \quad \text{So:}$$

$$1-(1-p_1) \ldots (1-p_n) \le p_1 + p_2 + \ldots + p_n - p_n(p_1 + p_2 + \ldots + p_{n-1})$$

Leaving out the last negative member, we get the n case.

This sum bound can be bounded with two further bounds:

$$p_1 + p_2 + \ldots + p_n < (1+p_1) \ldots (1+p_n) \le \frac{1}{(1-p_1)\ldots(1-p_n)}$$

The first < is trivial because by multiplying out the right side we get extra members.

The second ≤ follows member by member because: $(1+p)(1-p) = 1-p^2 \le 1$.

Infinite Event Sequence, Anti Achilles Paradox

Observing an infinite sequence of independent events, usually the question is how frequently they come true if they all have the same chance. Like repeating coin flips or dice throws and watching an outcome. This problem will be handled later but now we regard something much more general because don't assume the same chances rather being p_1 , p_2 , . .
Then the frequencies are not the concern, rather the more basic question whether they will come true infinite many times or merely finite times, that is stop occurring from a point.
We all feel that if something has a no matter how small chance and we keep tying it, then it must come true after a while. This of course means that it will come true infinitely.
Generalizing this to the chance from the actual event, then it would mean that trying different events that all have the same p chance will also have to come true and in fact infinitely.
Next then this also implies that if the events have not the same chance but these chances are all above a fix p value then again we must have infinite success.
And finally, even when they are not all above a p but infinite many of them are, then we can regard only these as a sequence and we get again infinite occurrence.
The negative of infinite many from a sequence being above any fix p , is that for any p value, the values will all stay under p from a point. This is what we call approaching zero or diminishing.
Thus, finite occurrence or stopping is only expectable if the chances are diminishing.
Amazingly this is not enough! The chances may diminish and yet keep coming true infinite many times. The crucial extra condition is based on a simple hidden feature, nothing to do with chances.
This hidden feature is the total sum of the values. More importantly its finiteness or infiniteness.
The Achilles Paradox was that infinite many points in time can approach a moment.
Of course, behind this lied the same in space that we can approach a point by going closer and closer. The vision behind this is simply that we move smaller and smaller distances.
This might give the impression that this adding up smaller and smaller values is actually a cause of the limit. But this is false! We can add diminishing values to our starting point and yet never reach a limit, rather go to infinity. We obviously must then go to infinity very slowly.
At first we can not even give such example because all our examples jump to fast motions.
The trick is to start with a fast motion, that is using fix steps first and then separate these steps into more and more and thus smaller and smaller ones. The most useful to start with $\frac{1}{2}$ steps that is with the infinite sum: $\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \; . \; . \; . \; . \; = \infty$.

Now cut the second member into two, the next into four then eight and so on:

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

We proved our point because the members are diminishing but we might feel a bit cheated by the repeating members.
So now lets increase our sum a tiny bit even further and thus still get ∞ as follows:
We increase the first $\frac{1}{4}$ to $\frac{1}{3}$ then the first $\frac{1}{8}$ to $\frac{1}{5}$ the second to $\frac{1}{6}$ the third to $\frac{1}{7}$.
As we see, the fourth can be left as $\frac{1}{8}$ and it will fit in exactly. The same trick can be used for the next $\frac{1}{16}$ group and increasing them starting to $\frac{1}{9}$ and so on will finish with leaving $\frac{1}{16}$.
So, these increases give exactly all the reciprocals and thus amazingly:

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

Fast And Slow Diminishing

Now we can precisely define the distinction among the diminishing p_1 , p_2 , . . . chances.
They are fast diminishing if $p_1 + p_2 + . . = L$ finite limit and they are slow diminishing if $p_1 + p_2 + . . = \infty$. The two grand claims are:

If the chances are fast diminishing then the events must stop.

If the events are independent and they do not diminish fast then they come true infinitely.

This sounds very physical but neither the assumed probability values nor the claimed stopping or continuing occurrence can be established experimentally.
A better meaning is only mathematical directly, with physicality hidden behind.
But it can't even be quite direct from our earlier facts because we only defined finite chance combinations and this relates to infinite. More importantly, it can't even be a simple limit chance of the beginnings because the stopping or continuing relates more to the end of the chance sequence. So we need two limit takings for the p_m , p_{m+1} , . . . , p_n segments. First to go to infinity with n and then with m. The meaning of this limit will be a shift from the definite claim of stop or continuing to a mere 0 or 1 chance value assigned to these claims. The actual physical interpretation of 0 as impossibility and 1 as certainty is then merely a subjective step.
As a detour, we have to realize that this interpretation is controversial. Indeed, the example we had with 0 chances was the dart throwing. But there, these 0 chanced individual landing cases at particular points are not impossible at all. We can lament about the vague plausibility that if someone would specify a particular point then it is impossible to land exactly there. We can also lament about an opposite direction that maybe these stoppings or not stoppings in sequences are not sharp actualities in a physical sense either.
All these relate to a deeper problem with randomness in general.
The sequential randomness will be handled later but the continuous one like at dart throwing, doesn't even have a theory yet. Before even returning to our problem, we should derive the above mentioned seemingly trivial physical claim that chances all above a fix p value must come true infinitely. The mathematical claim is merely that this infinity has a 1 chance. A finite occurrence means stopping and thus no occurrence from then on. The chance of this is the product of all the complement chances which are all under $1–p = q < 1$. But the product is then under the powers of q which must approach 0. So indeed the complement is trivially approaching 1.
Now back to our dilemma.

The chance of p_m or p_{m+1} or . . or p_n is: $\leq \; p_m + . . + p_n$.

Applying limit in n, the chance of p_m or p_{m+1} or . . $\leq \; p_m + p_{m+1} + . .$ too.

If $p_1 + p_2 + . . = L$ then $p_m + p_{m+1} + . . = L - (p_1 + . . + p_{m-1}) = d_m$.

So this d_m is the chance of the E_m event of at least one occurrence from m on.
It can be arbitrary small because $p_1 + . . + p_{m-1}$ approaches L.
The E event of having infinite many occurrences in the whole sequence implies every E_m .
Thus, the chance of E can not be bigger than the chance of any of the E_m that is d_m .
But since these d_m have arbitrary small values, the chance of E can only be 0.

Now we show the reverse for independent events.

If $p_1 + p_2 + . . = \infty$ then the chance of E is 1.

For this we re-write the earlier derived boundings of the chance sum.
First of all, we use it for p_m , p_{m+1} , . . . , p_n that is:

$$p_m + p_{m+1} + \, . \, . + p_n \;<\; (1+p_m) \, . \, . \, (1+p_n) \;\le\; \frac{1}{(1-p_m)\ldots(1-p_n)}$$

But we also regard the reciprocals of the three expressions, which means that:

$$(1-p_m) \, . \, . \, (1-p_n) \;\le\; \frac{1}{(1+p_m)\ldots(1+p_n)} \;<\; \frac{1}{p_m+\ldots+p_n}$$

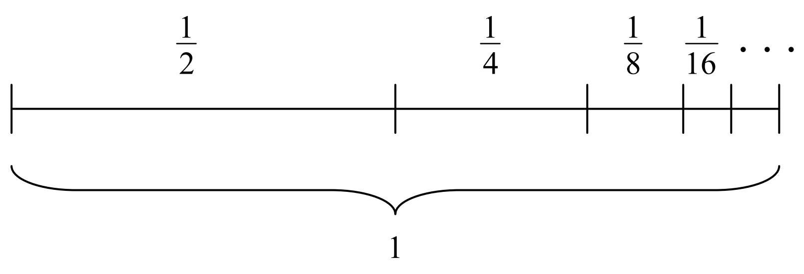
Now, if $p_1 + p_2 + \, . \, . = \infty$ then $p_m + \, . \, . + p_n$ can be arbitrary big for any fix m.

In particular, starting with $m = 1$ we can find an n that we have a sum of at least 2.
Then from this n as m we can find a new n that the sum is again at least 2, and so on.
Thus the consecutive segments will have $(1-p_m) \, . \, . \, (1-p_n)$ product values all under $\frac{1}{2}$.
Then the $1-(1-p_m) \, . \, . \, (1-p_n)$ chances of at least one occurrences in the segments will be all above $\frac{1}{2}$. But events above a fix chance must come about infinite many times with 1 chance.
This means at once 1 chance of infinite many occurrence in the whole sequence, that is of E.

Some Infinite Sums

The exact calculation of the $p_1 + p_2 + \, . \, . = L$ limits is a whole field of mathematics.
The most trivial case is what we already used to cover a point sequence with ε by:
$\frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \ . \ . \ . \ = \varepsilon$. The ε is of course is immaterial and the point is:

$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \, . \, . \, . = 1$ which is almost obvious from drawing a picture:



The more precise argument is also visible, namely that the leftover towards 1 after every beginning sum, is exactly the last member. In general: $\frac{1}{2} + \frac{1}{4} + \, . \, . \, . + \frac{1}{2^n} = 1 - \frac{1}{2^n}$

Indeed, it’s true for the start and inherits to new sums. Or directly with a little trick:

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

The next famous example is the reciprocals of the squares: $\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \, . \, . \, . = ?$

2^n is much bigger than n^2 if n is a big number. The only n where n^2 is bigger is at $n = 3$. And they only are equal at $n = 2$ and 4. So we would think that comparing with the previous sum, this will be much bigger maybe even infinite:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{64} + \frac{1}{128} + \ . \ . \ . \ = 1$$

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \frac{1}{49} + \ . \ . \ . \ = ?$$

Quite amazingly, we can show that the sum is still under 2 :

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \frac{1}{8^2} + \frac{1}{9^2} + \frac{1}{10^2} + \ . \ . \ .$$

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

$$\frac{1}{2^2} \qquad \frac{1}{4^2} \quad \frac{1}{4^2} \quad \frac{1}{4^2} \qquad \frac{1}{8^2} \quad \frac{1}{8^2}$$

$$1 + 2\cdot\frac{1}{2^2} + 4\cdot\frac{1}{4^2} + 8\cdot\frac{1}{8^2}$$

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

We can give an other proof:

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

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

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

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

The exact value of the sum is $\frac{\pi^2}{6}$ which took years for Euler to establish.

An other grand result of Euler was to show that summing the prime reciprocals we get ∞.

This shows that the primes are much more frequent than the squares, which is not surprising because between every pair of squares there are more and more primes. Amazingly, it is still not proven that there is definitely at least one prime in between.

Randomness, The Law Of Large Numbers

Now we can come to the mentioned much simpler situation when a fix p is repeated and we look at the possible outcome sequences denoted as 1 , 0 , 0 , 1 , 1 , 1 , 0 , 0 , 0 , 1 , 0 , 1 , . . .
We use the 1 for success and 0 for failure of the watched event. For example, with a dice usually the 6 is the desired outcome and the 1-s denote these. We expect one sixth of the outcomes to be this in the long run. So, in an N long beginning, the n number of 1-s is about one sixth of N. In other words, $\frac{n}{N}$ should approach $\frac{1}{6}$ or in general p.

The name of $\frac{n}{N}$ is relative frequency. Indeed, n is the frequency of the 1 successes and by dividing it with N we relativized it to N.

The heuristic importance of this so called Law Of Large Numbers is that it actually defines the probability so we don't have to assume it as a completely unexplained number attached to events.
The biggest champion of this approach to build a probability theory from these derived not merely assumed p values, was Von Mises. His unsuccessful approach lead to the opposite and accepted probability theory of Kolmogorov. But this doesn't mean that Kolmogorov didn't wreck his mind just as much about an alternative approach.

The fundamental problem in defining the probability from the $\frac{n}{N}$ limits is that this fact in itself is clearly not enough to claim that the outcomes were from the trials of such p chanced event.
The 1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 1 , . . . sequence has a perfect half frequency of successes and so obeys the $\frac{1}{2}$ chance associated with a coin toss. But it's too perfect! Such alternating outcome sequence is not random, it is a mathematically constructed sequence. Of course we don't need a perfect determination of the outcomes. It's enough if every tenth digit is a fix or even if these alternate in perfect order. Such subsets can at once defy randomness.

Von Mises basic idea was the recognition that requiring an obedience to the Law Of Large Numbers can actually detect even such sneaky imitations. Indeed, if we make every sixth outcome a 1 then it can not be a dice sequence, not only because we see this insane perfection, but because if we look at only every sixth positions outcomes, then they are all 1 and so the Law Of Large Numbers fails on this subset. We don't have one sixth success rather hundred percent.
So the rule seems to be simple: Every subset must obey the same p probability limit.
But we just entered a world of pain.

There are two instant problems. One shows that this rule is too little and one that it is too much.
So why is it too little? Well, because not every rule is prescribed on fix positions but can mean a strangeness defined by the outcomes themselves. If for example we see that after every second 1 outcome we always have three 0-s then this is not normal. Okay you may say, then lets allow subsets selected by these self outcomes too and we then catch the fix digits again. But this makes the other problem even more relevant. Namely, that even in a perfectly random sequence we can select the certain outcomes and so they obviously fail the chance frequencies.

So, the usable observational subsets must be selected by some methods that are free enough but can not rely on the outcomes themselves. The crucial compromise seems to be quite easy.
The past can not influence the future! This law sounds plausible and yet we all have a tendency to feel contrary to it. Namely, after twenty heads in a row we feel that a tale must be more likely.
I made some research about this false belief and not just to justify my own stupidity. At any rate, if we get over this false plausibility and accept that the past, that is a whole beginning of outcomes can not influence even the next outcome, then the observational subsets are easy. They should be such next places after any prescribed beginnings, that is beginnings obeying some features.

These beginning features then can include merely the length of the beginnings, that is in fact giving the positions of the next places but can also depend on the outcomes in the beginning.
Total success! Unfortunately, we didn't define exactly what such beginning feature can be.
Subjectively it's obvious that it must come from the finite long binary sequences as beginnings.
But we don't have a method which guarantees that we indeed selected certain such beginnings to tell the next positions and not in reverse. Indeed, the beginnings before all the 1 outcomes in a particular sequence is itself a feature and the positions defined by it will be exactly the 1 outcomes. So, observe that the problem is not just the reversed defining of the beginning from the next outcome, rather the application of this to a particular abstract sequence. Or to be precise, the real problem is a mixture of the self-determination combined with abstract random case.

Church was the first who realized that machines are a magic wand that can avoid this catch twenty two. They are ignorant of the abstractions of mathematics and this allows them to avoid the above reverse determination for particular random sequences too.
So, a machine that accepts, that is stops from certain binary inputs should be used!
The set of these recognized beginnings can even be visualized as circling some members from the list of all possible beginnings:

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

Of course, the circlings, that is recognitions don't have to be in increasing order. A short might only be recognized arbitrary late by the machine.
Now, looking at any given fix binary infinite sequence, if we spot some beginnings there, being one from the circled ones then the next place after that in the given binary sequence is an observational place. The set of all these, that is after all the spottable beginnings is the observational subset determined by our machine. We can regard other machines for the same given binary sequence and our claim is that if these machine selection places all have the same p limits then the given sequence was random and came from repeating a p probability event.

The first thing to emphasize is that above the "looking at" meant looking carefully in the whole given sequence as if it were given in full, not merely watching it unfold as outcomes in time order.
So, these observational place selections are wider than some forward going prediction places.
Such real life prediction couldn't wait for a short beginning to be recognized. But this is good because we went broader, so made sure that even such unrealistic selection places are allowed.
In fact as an instant idea to regard more realistic selection places could be to require that the machine is not merely a recognizer but a decider. Which of course as we explained can only mean realistically having two machines in one box, that is recognize the complement beginnings too.
Then we still need arbitrary long time to make the selections but at least we can go in time order and check out all appearing beginnings in the given binary sequence. At any rate, allowing the wider range of all single machines is better as start, bringing out the heuristic role of machines.
We might even think that this role is to avoid the application of individual sequences in general and that's why they avoid the random ones trivially. But this is not true! The machine role is deeper! A machine can simulate ruled sequences and then accept some beginnings according to outcomes that are after the beginning. But this is okay. Indeed, a random sequence can not have the same beginning dependence from the later outcomes infinite many times.

So the point is that machines can avoid random sequences. And thus, we mathematicians who can't avoid them can use the machines to separate them. But how we use them for this separation, is still our choice. In this Church idea the separation was based on accepting Von Mises beliefs:
If for all machines the observational subsets have same p limits then this p is the hidden probability hiding behind the particular sequence and this sequence is random.

Or in negative, if for a binary sequence there are two machines that the places after their recognized beginnings are two sets of locations where the outcomes don't approach the same p, or one or both of them give places that the outcomes don't even approach any p at all, then the sequence is not random. One machine can already define a strangeness as observational places that have no limit. Two machines can define strangeness of having different limits.

This machine approach makes sense from an other practical angle too. Whatever we can describe as a beginning feature, if it is really an exact feature, it should be translatable into a mechanical process of recognizing such beginnings.
Also observe that the usage of beginnings to avoid self outcome dependence, magically coincided with the machines using finite inputs. It is too beautiful not to be perfect.
I have to be a spoil sport and reveal that it wasn't perfect.

The problem was not Church's idea to use machines, rather that he missed a heuristic new vision how to use them and instead accepted Von Mises' false belief.
But then the strange fact is that this missed vision remained under ground for fifty years.
In fact never resurfaced perfectly.

For me it is not merely strange but spooky because more than forty years ago I had that vision and few years ago I also contacted the person, Rod Downey who is the closest to be called the definer of that vision. At the time I contacted him, I wasn't aware that he was the one who made the final definition. Sadly, he is not aware of the vision behind his actual definition either.
So, a lot of missed awareness is going around, which is really the point of this book and my whole philosophy. A quite opposite kind of mathematician I wanted to contact is Jean Yves Girard.
He is aware of something that he even made the title of his book "The Blind Spot". But then in the details he steps into the traps of Formalism so easily that it destroys his correct awareness.
Back in high school when the three visions of randomness, non standard calculus and forcing came to me, I wasn't aware of anything.
I didn't speak a word English but I still could have made better enquiries about what was going on in the world. Instead I behaved like a hermit and wanted to discover everything by myself.
Today I know exactly how wise my instinctive ignorance toward the world of professional mathematicians was. Even by the unavoidable educational influences I sucked in enough Formalism that took decades to get rid of.
Only now as an old man can I again be like I was meant to be. The wasted times between, is what we call a life. But what really counts is what you see, what you are aware of at any moment.
I said that back then I wasn't aware of anything. Most importantly, I wasn't aware of why these magical abstract concepts defying every law or logic of matter just appeared in my head.
Even when I became a convinced idealist and I understood the whole point of philosophy from Hegel's approach that ignored the crucial role of mathematics, even when I dropped acid and understood Timothy Leary's death sentence of the establishment that ignored again mathematics, and finally, even when I realized that mathematics is merely an eternal narrow peek into a bigger but time filtered universe of understandings, I still didn't face those earliest encounters with the other side. So, to be beyond vanity is not enough. You are given and you owe. It's that simple!
And I'll try to show. I will explain the simple missed meaning of randomness that you will not find in any of the thousands of abstract articles floating in the junkyard, we call the internet.

The first step is to explain, why the Church approach was faulty.
As I told, it was not his ingenious idea of using machines rather blindly believing Von Mises.
The fundamental line of attack against Von Mises' obsession with the $\frac{n}{N}$ relative frequencies, that is the Law Of Large Numbers, came from a very simple fact relating to something that we actually already mentioned. It was Borel's Monkey, typing down the bible by pure chance.

The Law Of Big Numbers

Unlike the Law Of Large Numbers that seemingly claims infinite tendencies, this law that I call the Law Of Big Numbers only claims that whatever finite situation has a chance, must come true if we try it enough times. So, even to have a million heads in a row is possible. All we have to do is restart and repeat a million flips. Or since we feel that time is just like space, we can simplify the situation and flip a million coins simultaneously as one trial and then simply repeat this.
A missing side of this healthy intuition behind this Law Of Big Numbers is that though we know that these situations must be possible, we have a pretty lousy estimate of how rare they can be sometimes. But this just follows from our general lack of conceiving big numbers.
Indeed, the story goes that the creator of the chess game was offered any present from his king and he humbly asked for some rice. Namely one grain for the first corner place on the chess board two for the next and so on doubling them in number. Having sixty four squares, this means 2^{64} many grains for the last square, which is more in volume than the earth. This is also the number of trials that we would need in average to get sixty four heads in a row. This leap to cosmic times is not following from our intuitions. 2^{10} is only 1024 and that many trials can be done with ten coins flipped simultaneously in an hour.
The chess example shows that Borel's Monkey typing down the bible would require even more cosmic times because we have more symbols than two and have a much longer text than sixty four symbols. Still, the Law Of Big Numbers apply. A special meaning back to coin flips means that arbitrary long full heads or tales can be obtained by long enough trials and a single infinite sequence always can be regarded as a sequence of these too. This means that we have longer and longer full consecutive segments of heads and tales. So the tendency of the heads and tales "evening out" is a false vision. The Law Of Large Numbers is not an intuitive claim. It does claim something true but that thing is not an elemental thing. It is a complicated fact and we simply use abstraction to cover this up. This is the simple root of why Von Mises was betting on the wrong horse and Church failed by following him. Luckily, there is a perfect intuitive vision that goes behind the Law Of Large Numbers. But first we have to show the thin ice that sank Church's approach.
The difference of the occurring heads and tales in an N long beginning is clearly:
$n-(N-n)=2n-N$ or $(N-n)-n=N-2n$ depending on what we regard as the chosen successful outcome the head or the tale and which is more.
This difference should pick up arbitrary big positive and negative values.
This would follow immediately from our used arguments that showed how arbitrary long heads and tales must occur, if we knew that there are infinite many beginnings where the heads and tales are exactly equal. Indeed, after these equalizations, we can place arbitrary long windows too and we will get infinite many same segments and thus arbitrary big oscillations too.
Quite amazingly, we can not guarantee the seemingly easy part, the infinite many equalizations by Church's method.
We can manufacture a sequence, that will have all of its observational subsequences with $\frac{n}{N}$ approaching $\frac{1}{2}$, and thus qualify as random, yet achieve this approach from always being under or always being above $\frac{1}{2}$.
The idea behind this construction is to list all possible beginning recognizing methods and make sure that they are all under or over defined in relative frequency.

Behind The Law Of Large Numbers

As I said, I regard this law as an abstract truth not a plausible principle.
The easiest way to see my point is first to realize that this law actually is a totally opposite claim than it appears to be. Indeed, $\frac{n}{N} \rightarrow p$ seems like a tendency toward infinity.

Of course we can simplify this to $|\frac{n}{N} - p| \rightarrow 0$ with using the absolute value of the difference of the relative frequency and the probability. So the Law Of Large Numbers is actually a claim of diminishing. We dealt with diminishings earlier but now we still not use all those results yet.
Merely the simple fact that it meant staying under arbitrary small ε value from a point on.
So this means that for any ε there is some M number, that if $N > M$ then $|\frac{n}{N} - p| < \varepsilon$.
Of course, to have such M number, simply means that there can only be finite many N values where $|\frac{n}{N} - p| \geq \varepsilon$. Or even simpler: $|\frac{n}{N} - p| \geq \varepsilon$, must stop.
Observe that p is an assumed fix value. The n number of 1-s is determined in every sequence by choosing the N length beginning. So $|\frac{n}{N} - p| \geq \varepsilon$ is simply a beginning property of outcome sequences for every given ε. We claim that these properties always stop in sequences that correspond to outcomes of a p probability event.
Quite oppositely, the infinite occurrence of any of these properties are actually strangenesses, tell tale signs of a strange sequence. This makes perfect sense. A strangeness feels like something occurring infinite many times that shouldn't by mere chance. Indeed, a finite occurrence can merely be a coincidence for any property by the Law Of Big Numbers.
So, that's what we should strive for. To find all those beginning properties that shouldn't occur infinitely. Then we can define a sequence being strange by having any strange property that occurs for infinite many beginnings and being random as the opposite, that is all strange beginning properties stopping in the sequence.

The real beauty is that we can easily define these strange beginning properties in general.
It is just as heuristic as the foundation of logic was through the Taboo Avoidance as matrix concretizations. There we had the three different finite restrictions. Here we have the same again.
The first finite versus infinite is the use of finite or infinite occurrence as mere coincidence or strangeness. The second is using Church's heuristic idea. The beginning properties should be sets recognized by machines. This is the qualitative finiteness requirement about beginning properties.
But this is not enough. We need a new quantitative finiteness requirement too. This is obvious by the simple fact that we can list all beginnings easily:

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

These regarded as a strangeness would make every sequence be strange because infinite many of these must occur in every sequence. We can not allow to collect this many beginnings by a machine. So we must restrict the machines externally by something that can not be told by them as machines rather from the set they collect. These collected sets are all either finite sets or usually infinite sequences so it's not some smaller infinity we need. An obvious similarity shows with the earlier infinity versus measure. We need some kind of measure restriction.
And indeed, this was the historical road to this quantitative finiteness restriction.
But we can avoid that. I did when I was sixteen and knew nothing about machines or measures.
So I struggled with the more important qualitative finiteness and then gave up.

Beginnings That Must Stop, The Fast Diminishing Vision, Stop Conservation

My short obsession with randomness started as a Galileo kind of wondering away, from the Goldbach Conjecture. Usually this is formulated by the claim that every even number above 2 is the sum of two primes: $2n = p + q$. This of course means also that $n = \frac{p+q}{2}$.

Both forms could allow that $p = q$. And indeed, the first two even numbers above 2, that is 4 and 6 need this possibility because $4 = 2 + 2$ and $6 = 3 + 3$ are the only ways. But from then on, we always have different p , q too. For example $8 = 3 + 5$, $10 = 3 + 7$ and so on.
So a much better way to formulate the claim is that for every n number above 3 we have a p prime under n so that there is a q prime above n that n is the average or the middle number of p and q. To put it even better, if we mirror p to n then it is again a prime.
The opposite of this would be that if we mirror all the primes under n they would never land on a prime. Regarding the primes as successful occurrences with 1 and the composites as 0, we get a sequence. The mirroring of all the 1-s up to a point and not coinciding with the later 1-s could be called an anti symmetry of the twice long beginning. A symmetry would be that all the 1-s are mirrored, a repetition could mean that instead of mirroring we shift the first half and all 1-s move into 1-s, and an anti repetition could mean that they never go onto 1-s.
All these beginning features are very strong requirements, obvious strangenesses and so they must stop in a random sequence. The primes of course are not random and yet they obey these stoppings. Which probably means that these beginning properties are not forced by the primeness and so they behave just as they were unrelated to the primeness that is regard primeness as random. My interest at once shifted to the question of whether these stoppings in real random sequences are indeed a plausibility. So, I tested all my family members with the simplest question whether in a coin flip sequence a perfect repetition of beginnings could happen infinite many times or not. They all instantly felt it to be impossible. The real mystery in this is that here we have a new second plausibility beyond the Law Of Big Numbers. I rejected the phony plausibility of the Law Of Large Numbers very soon as a tendency, but now seemingly we must have some hidden tendency behind this intuition of stopping because many infinite occurrences are obvious by the Law Of Big Numbers. This mystery is even deeper after we realize that this intuition can not be a simple felt tendency of diminishing. Or it shouldn't be because there are diminishing chances that keep on re-occurring. The earlier section of Fast And Slow Diminishing and the one before about the Anti Achilles Paradox, are the clarifications of this but not with beginnings in a fix sequence rather in general.
So how could the special case about beginnings hide clearer plausibilities?
In fact, I wasn't sure of two important questions. The simpler is whether slowly diminishing beginning properties would plausibly be differentiated from the fast and judged to be continuing rather than stopping.
The second is whether a crucial mistake in all these thoughts have a role. This crucial mistake is that beginnings are not independent in a fix sequence. They have to continue each other. The clear results of the two kinds of diminishings were all for independent events and so they can not be used for beginnings in a fix sequence. Only for beginnings as restarted new trials.
Actually, there are three different possible connection of segments to a sequence:
The most obvious is the continuations, that is the segments all regarded as beginnings. That was the sense that I tested my subjects and they seemed to have a very consistent plausibility. But this is a mystery because this continuation is a mixture of two factors. The feeling about the chance of the property but also a feeling about this being burdened by that it has to be continuing.
The second sequencing of segments is the very plausible Law Of Big Numbers applied as a single sequence. Exactly as we explained, that is as fix windows infinite many times. This at once shows that arbitrary long self repeating segment will occur just as any other wanted segment. The Bible as a segment will also occur and so Borel's monkey will eventually type it down. But the paradoxicalness of this shows that this line was an intellectual or rational line, not directly visual.
The third way to make a sequence out of segments is even more "rational seeing". Namely, we can regard not fix, rather increasing windows. So we flip one coin, then we flip two, then three

and so on. Will a property continue or stop? Here the plausibility is missing. So quite amazingly, the infinite occurrence of self repeating segments at the even windows will not be denied by most people. I don't feel it absurd either, just as I don't feel absurd to wait a few rolls at roulette after a long segment of reds, to get a red again. These black holes in our plausibilities mean something but I have no clue what. The rational line of course can penetrate and we finally "see" that indeed the red after hundred red is just as possible. The line involves looking at the previous equal windows. And similarly the stopping of the repeated segments can be "seen" by an even more complicated tour of looking at the whole ends of the sequence as we will show it later. But then why was the original stopping of the repeating beginnings so evidently plausible? I don't know!

We can't just say that beginnings are more plausible than segments! For example, the beginning properties that claim some end segments are all trivially infinite to our intuition. But also some non segmental ones can be too! Like the half half chance of having more 0-s or 1-s. That's why the failure of Church's approach was so trivial. A crucial case would be to find a diminishing one that is plausibly continuing too. This could at least reveal whether the slow or fast diminishing has a plausibility. The closest to a "plausible" one I could find is the property of a beginning to have a last outcome repeated at the end more times than anywhere before in the beginning. We know that such beginnings must occur in a sequence infinitely because we know that there will be longer and longer full 0-s and 1-s. And these are indeed diminishing in chance because otherwise the Law Of Large Numbers were false. Which I denied as plausibility and here it was not a matter either.

The even more sad fact is that the crucial quantitative finiteness of the strangenesses will come from this non intuitive distinction of fast versus slow diminishing and will completely ignore the seemingly much more intuitive restriction of continuation.

At first this sounds even contradictory and we might think that we can demonstrate an effective beginning collection that is claimed to be a strangeness because the set is fast diminishing and yet some random sequence sneaks through it, that is doesn't stop because the beginnings help the continuations. But that's impossible! The beginning influences can only be opposite, that is making some continuations impossible. And this has nothing to do with diminishings. Even positive chance can stop due to the continuation. The obvious case is to make a claim of some fix beginnings to be or not to be as sub beginnings. Simplest is to claim that a beginning starts as 1. This has fifty fifty chance but once it happens it will inherit and stop the fifty chanced 0 start.

So, beginnings do influence the continuations, very drastically. In a sense, only drastically. So the stopping of the fast diminishing properties is conserved from the consecutive segment trials to the real continuing ones. This is the real seed of our whole randomness concept. We simply apply the law of the stopping of diminishing independent events to the non independent beginnings.

Compared with the totally Formalist approach how new math is handled today, my approach is a didactical heaven. But as I said, I believe in a perfect didactical logic of the future and compared to that I'm in constant darkness. At least I draw attention to this and even try to paint a picture where the future lies. The three finiteness in Logic was also imperfect. Much better than some idiotic ad hoc start from a set of rules as all text books go. I showed that such rule systems are all merely consequences of the matrix concretizations. But the fact that the non the existence of finite contradictions leads to models, that is the Completeness Theorem is still not a plausibility.

Similarly, the strangenesses as infinite continuations in both qualitatively finite that is machine generated and quantitatively finite, that is fast diminishing beginning collections is not quite plausible. These two simple and beautiful principles will connect when a new better Set Theory will be found. Then the truth becomes pure understanding.

Right now we have to finish our program and precisely define the quantitative finiteness because the fast diminishing is merely a subjective vision.

Remember that for diminishing numbers we already established what this fast diminishing means. Namely that the total is a finite sum and not infinite. But for beginning collections we have to define some chance values inside that could mean these. The real beauty would be if the mentioned finite total of these chances would be the quantitative finiteness. The strange consequence then would be that chances summed up and being bigger than 1 which is the full probability of certainty, could still have a new meaning. This will exactly be the case.

Length Partitioning

A set of beginnings will be denoted as B. This can be collected by a machine and then usually I call it as a pool. The strange sequences must pick continuing beginnings from the pool if they can. If there is no such possible continuation at all then I call the pool void. Finite pools are all such. But infinite pools can be void too and being void is not the only reason a pool can be useless as strangeness. Namely, quite oppositely, it can be too big like the set of all beginnings. So we have to narrow these by some probability meaning. This comes through the length partitioning.

A partitioning of B is $B = B_1 \cup B_2 \cup \ . \ . \ .$ The $\cup$ means union or combining the members. We also assume that any two of these members are disjoint, that is have no common elements. The simplest partitioning is the length partitioning for which we use the special notation of upper indexes. So B^1 contains 0 or 1 or both, B^2 contains some elements of $\{00 , 01 , 10 , 11\}$, then B^3 contains some elements of $\{000 , 001 , 010 , 011 , 100 , 101 , 110 , 111\}$ and so on. These combine to B and so B itself determines these. They are simply the listing of the longer and longer beginnings in B. Obviously, there are 2^n many n long beginnings altogether.
Now we can come to the crucial probability meaning of this length partitioning:
The length of a b beginning is denoted as $< b >$. For example: $< 10110 > = 5$.

At a coin flip, that is with $p = \frac{1}{2}$, the different b beginnings with same n lengths have same chances, namely: $| b | = \frac{1}{2^{<b>}} = \frac{1}{2^n}$. The chance value of B^n is $| B^n | = \frac{\text{num}\, B^n}{2^n}$,
that is the number of elements in B^n divided by the number of all n long beginnings.
This is the chance of obtaining a B^n element if we try n long beginnings repeatedly.
An even better visual form of $| b |$ is obtained by regarding b's binary form and then simply replacing all digits by 0 except the last with 1 and finally put a decimal point in front of this:
$| 10110 | = .00001$ but this is meant now in binary "decimal" value which of course is $\frac{1}{2^5}$.
Then $| B^n |$ is also regardable as the sum of these $.0...01$ chance binaries.

For the general p case all these aboves are not true! The chances of same long beginnings can be different and so we have to look at what elements are in B^n not merely their number.
For example at $n = 5$ then the 10110 and 00110 elements have different chances.
The first has $p\,(1–p)\,p\,p\,(1–p) = p^3 (1–p)^2$ the second has $(1–p)\,(1–p)\,p\,p\,(1–p) = p^2 (1–p)^3$.
At any rate, we can still add up these chances in B^n and obtain $| B^n |$.

The beauty is that the $| B^1 |$, $| B^2 |$, ... chance value sequence being fast diminishing means $| B^1 | + | B^2 | + \ . \ . \ . = L$. But each member is a sum of individual $| b |$ beginning chances. So the L limit is actually the sum of all beginning chances for the b beginnings in B.
In symbols : $L = < B > = \sum_{b \in B} | b |$. I use this $< B >$ notation instead of $| B |$ to emphasize that this sum of chances is not an actual chance itself.
So $< B > =$ finite is our crucial third or quantitative finiteness.
Just to remind you, the second qualitative finiteness is the machine usage to define B pools and the first finiteness is the heuristic meaning of random sequences occurring only finitely, that is stopping in any B pool that is both quantitatively and qualitatively finite, that is, is a strangeness.
If the $< B > = \sum_{b \in B} | b |$ total sum of chances is ∞ , then the pool will not define a strangeness.
Then the random sequences don't have to stop in the pool, they can continue infinitely.
This infinite occurrence or the "obeying" of the B set is not a strangeness then, rather natural.

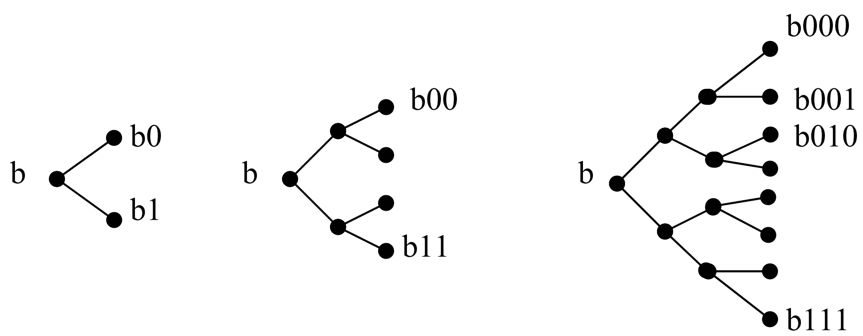
A Deep Strangeness Of Strangenesses

From the machine that collects B, we can not establish directly if < B > is finite or not.
Empirically of course, using the machine to collect more and more beginnings in B and adding up their chance values, we get an approximation as an increasing sum.
If < B > is finite, we eventually would get longer and longer beginnings of its decimal form too.
Unfortunately, this “eventually” is purely theoretical because we never know if they are already established or not. For example, we see a total of 23 **.** 506 slowly increasing. This still allows that we will get 374 **.** 45 and so the earlier digits were totally useless. We might even climb to infinity.
This is puzzling enough, but now suppose we knew somehow that the exact beginning of the sum is indeed 374 **.** 450932. This would give us an amazing knowledge because then after we climbed in our empirical sum to some beginnings of this known beginning itself, we could establish the non recognizabilities of short enough beginnings. Indeed if they were not recognized earlier that is already added to the empirical sum then they being recognized would tip the sum over the known value. So we could recognize a lot of non recognizabilities. If all digits of < B > are known then gradually we can establish all unrecognizable beginnings, the whole complement of B.
Of course B itself can be a totally derivable set, that is have an effective complement and then this is not surprising at all. If B is a typical effective set with a non effective complement then obtaining the complement from < B > is not that surprising either because < B > is not effective.
It is empirically concrete, namely it is machine determined but not machine recognizable.
This sounds a bit paradoxical but if we think about it then all unrecognizable complements are machine determined too. Here the surprise is merely that this < B > value is more useful then simply a complement. Indeed, it can give the unrecognizable beginnings in increasing order.
After we calm down and accept all this as a normal feature of effectivities, comes a real shocker.
What if B is collected by a universal machine? Well, as a trivial consequence, the complement then has to be non effective but that’s not all! The inputs are now programs for all other machines.
So the longer and longer unrecognizable beginnings are now bigger and bigger machines’ unrecognizabilities. So the < B > value actually gives all possible infinite running situations of all possible effectivities. Most famous claims of mathematics are such infinite runnings.
Namely, having no counter examples for some claimed universal situations. For example, the Goldbach Conjecture claims the situation that mirroring the primes under a number bigger than 3 always hits a bigger prime. A machine can verify this for any n number easily. Thus, starting from n = 4 and trying all numbers one by one, the running for ever means the truth of the claim.
So, the < B > number doesn’t give the proof of all these claims but verifies them all.
If the Goldbach Conjecture were false then we would encounter some huge number where the smaller primes mirrored would fall onto all composites. This might even seem plausible by knowing that the primes become very rare. But it’s false because the amount we mirror is increasing. So the true tendency is that there are more and more hits onto bigger primes among the mirrored smaller primes. But the idea of possible huge counter examples shows that recognizings become very slow. We have to wait for eons to reach even the first few known digits of < B >.
This is a practical problem even if we knew some initial digits from some hidden source.
The more important question is whether we can derive some initial digits by analyzing the machine? Individual infinite runnings can always be established from lucky observations about a machine. Establishing all infinite runnings up to some program length still doesn’t give the digits up to that length because the longer ones influence the total. But they could give some digits.
The anti diagonal results of effectivities are merely forbidding universal methods to establish all runnings. But here, where the knowledge of the initial digits is spitting out all shorter unreconizabilities, we should get some stronger impossibilities beyond the world of Turing.
Solomonoff, Kolmogorov and Chaitin were the explorers of this “beyond”, which is information.
Chaitin proved that from the machine that collects B, we can only derive a fix many digits.
This first seems absurd because a finite set is always derivable. So if we know any number of initial digits from a hidden source then we can build a machine that uses that as data and so pretends to predict the digits from the machine. So, seemingly we only have the triviality that one machine can tell only finite many digits. But Chaitin turned this triviality into something deeper.

Namely, by telling from the complexity of that machine the limit of the predictable digits.
But not quite! Indeed, this still depends on the original machine too that collects B. To analyze such machines, the other one should be an axiom system like Set Theory. The really weird thing is that then the collecting machine can influence the predictable digits to be extremely low. Solovay proved that for some machines that collect a B, Set Theory can not tell any digits of $< B >$.
To interpret this as a total restriction on derivabilities in general is false. It reveals that the initial digit predictions of $< B >$ is simply a special forbidden zone. Like the light speed limit of Relativity. An abstract restriction, that can be still interpreted in many ways and even empirically doesn't define an obvious field of impossibilities. Here too, Chaitin and his admirers exaggerate. The limit of an axiom system like Set Theory can not be purely measured by information contents. But his enemies claiming that it's all "smoke and mirrors" are also wrong.
Information is a limitation and implies incompleteness.
So Gödel avoiding to meet Chaitin was merely a sign of what we already know. Gödel was a psychopath. Chaitin not getting a Fields Medal is again a sign of what we know. Academia is insane too. Rice Theorem was a bigger result and still remains unrecognized.

Cover, The Kolmogorov Road, Concrete Sets Of Random Sequences

B^{∞} will mean the set of all sequences that are continuations of any beginnings in B.
A more visual saying of this is that B^{∞} are the sequences "covered" by B. Then this covering vision can be used for beginning sets too, namely a C covering B simply means $C^{\infty} \supseteq B^{\infty}$.
If the B set is a single b beginning that is $\{b\}$ then we simply write b^{∞} instead of $\{b\}^{\infty}$.
A b beginning always covers a c continuation of it, that is $b^{\infty} \supseteq c^{\infty}$, but continuations can cover a b too. For example: $\{00 , 01\}^{\infty} \supseteq 0^{\infty}$ namely equality stands here.
The deeper vision is that the infinite sequences are points and the beginnings are intervals. Indeed, the $[0 , 1]$ interval can be regarded as the points of the binary "decimals" that is as repeated halvings. The 10110 beginning is a fifth deep halving interval found by first regarding the second half of $[0 , 1]$. Then the first in that, then twice the second halves and finally again a first half.
The crucial step between sequences as points and the beginnings as intervals is the beginning sets. These are interval sets that can cover complicated sets of points, that is sequences.
I call this vision the Kolmogorov Road, that literally leads from the beginnings, through the infinite sequences onto the unit interval. He used this vision before randomness was defined for a quite opposite goal, namely to avoid the concept of randomness and regard probabilities as measures of point sets. But measures have problems and we'll come back to this later.
The most trivial visual assumption behind a cover is that its size can't be smaller than the size of what is covered. But size can't be simply regarded as the total length $< B >$ because these lengths overlap. So to claim things we should regard non overlapping intervals. We'll come to this.
But actually we can claim an obvious thing without this. Namely, the covering set C doesn't have to be assumed non overlapping. If it isn't then its real size is still at most its total $< C >$. So this is still an upper bound. On the other hand a covered B where the non overlapping is trivial, is a single b beginning. So the trivial claim is this:
If a C beginning set covers b that is $C^{\infty} \supseteq b^{\infty}$ then $< C > \geq | b |$.
Surprisingly, the proof is not that trivial. Lets regard all possible continuations of b :



Since every sequence continuing b is a path in these continuations, these paths must all have a c beginning lying in C too. If any of these is b itself or a sub beginning of b then trivially: $< C > \;\geq\; | c | \;\geq\; | b |$. If not then in every path going from b we have a beginning that lies in C and in the groups: { b0 , b1 } , { b00 , b01 , b10 , b11 } , { b000 , . . . } , . . .
Let the number of these in each group be m_1 , m_2 , . . . allowing of course them to be 0.

The total chance of these is $| b | \left(\frac{m_1}{2} + \frac{m_2}{4} + \; . \; . \; . \right)$ and so enough to show that:

$\frac{m_1}{2} + \frac{m_2}{4} + \; . \; . \; . \;\geq\; 1$. This seems quite unbelievable, knowing that we can have 0-s.

First of all, if finite many c_1 , c_2 , . . . , c_n from the C set already cover the b continuations then our claim follows easily. Indeed, let c_n be a longest among these and replace all shorter ones by this long continuations. The fraction values don't change. Also, we have to get all this long beginnings to be covered and so if they are without any repeats they still give 1 as total.
Now comes the surprise, that such finite cover always exists. Indeed, suppose it wouldn't. Then among the b0 or b1 continuations also at least one of them couldn't have a finite cover. Choosing this, we can again regard the two directions and chose one that doesn't have. And so on, we get an actual sequence that for no beginning of it had a finite cover from C. But that's impossible because this sequence had to have a beginning in C which was itself such cover set.

Now we can come to the crucial use of the cover vision for us, that is for strangenesses.
Indeed, the $< B >$ = finite quantitative condition means that in $| B^1 | + | B^2 | + \; . \; . \; .$ the more and more initial members approach the finite value and thus the rest, that is the "end sum" must approach 0. But this means that infinite occurrence from B, that is the strange sequences are "covered", by arbitrary small intervals. This instantly gives a way to see that all the strange sequences together are also arbitrary small and thus the rest, that is the random sequences are arbitrary close to 1 in size.
The extra assumption we need is that all B strangeness pools can be listed as: B_1 , B_2 , . . .
Then for any small ε we have: $\frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \; . \; . \; . \; = \varepsilon$. So:
We go in the $B_1 = B_1{}^1 \cup B_1{}^2 \cup \; . \; . \; .$ length partition of B_1 far enough to be under $\frac{\varepsilon}{2}$ in the end sum of the chances. Excluding these, we have $1 - \frac{\varepsilon}{2}$ chances of sequences that are definitely failing the B_1 strangeness. Then we go in B_2 far enough to be under $\frac{\varepsilon}{4}$ and exclude these too. And so on, gradually we define a set of sequences that avoids all strangenesses and has:
$1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{4} - \; . \; . \; . \; = 1 - \varepsilon$ chance.

Now some remarks about this assumed listing of the strangenesses:
The second qualitative finiteness or the machine collection is again arguable whether recognition or decision should be used. Of course, decision actually means total derivability, that is the usage of two machines. We already mentioned this at Church's original idea. The derivable sets are effectively listable because the machines are generable by their rule systems in a fix framework of effectivity. Like say binary Turing machines. These machines as recognizers give the sets. The totally derivable sets are also listable by unnatural parameters but not by the machine pairs that define them. This as a negative argument against using totally derivable B beginning sets is irrelevant though. The reason is simply that an effective listing of all possible B sets is useless anyway. Indeed, we need the quantitative finiteness $< B >$ = finite. But from the machines we can not tell this. Still, these strangenesses are a sequence of sets even if they can not be given effectively. That's why I used the word "concrete" in the title.
We can go further and chose some random sequences from these concrete sets and then we can get concrete individual random sequences. The most famous such case is Chaitin's omega

numbers. We already mentioned that the $< B >$ value is predictable only for a given few digits. He actually proved that this value is random too. Of course it is a decimal not a binary sequence. We'll come to his method to get concrete binary sequences instead of decimals.
These usages of the word "concrete" are simply what we earlier called explicit.
If we make theories of randomness then we always will have concrete random sequences.
But it is also a natural intuition that if something is given concretely then it shouldn't be random.
So, this initiates a new theory that tries to avoid the concrete cases. Of course unsuccessfully.

Non Continuing Beginning Sets, The Perfect Total

What made our argument above easy, was that the end members together were arbitrary small in the length partitionings. They covered the strange sequences. In fact, many members covered same ones repeatedly. But this didn't matter because it made the covered point set even smaller.
If the total of the covers is not diminishing then such overlapping, that is multiple cover, makes the total only an upper bound but not a reliable measure of the covered point set. But in our case the intervals are halvings so overlapping is actually being sub intervals. For beginnings this simply means continuation. For our whole goal of B beginning sets as strangenesses such continuability is the very essence. The complete opposite that is having no continuing pairs in B at all, implies what we called void, when no sequence can come about from B.
And yet now we examine such non continuing beginning sets for two reasons:
They will be used to get two other special exponents, 0 and bracketed (n) ones.
These will lead to a much better perfect partitioning. Also, these non continuing sets were what Chaitin used to get his concrete random numbers. So why are these non continuing sets special?

The fundamental fact is that their total is a correct measure of the sequences covered by them.
So, now we can give the general version of our earlier result:
The total of a cover is bigger than the total of a covered non continuing set:

If B is non continuing and C covers it, that is $C^{\infty} \supseteq B^{\infty}$ then $< C > \; \geq \; < B >$.

We regard the longer and longer $b_1 \, , \, b_2 \, , \ldots$ elements of B and will replace the C elements with longer and longer continuations too. The continuations of $b_1 \, , \, b_2 \, , \ldots$ thus form disjoint $C_1 \, , \, C_2 \, , \ldots$ sets and for each we can apply our earlier result so:
$< C > \; \geq \; < C_1 > + < C_2 > + \; . \; . \; . \; \geq \; | \, b_1 | + | \, b_2 | + \; . \; . \; . \; = \; < B >.$
But here the first $\geq$ is not obvious. To see that, we have to replace the $b_1 \, , \, b_2 \, , \ldots$ too with longer and longer versions. Then more and more of the disjoint sets are in same length groups.
This at once proves as a side result that they have maximum 1 value and so also $< B > \; \leq \; 1$.
This is very natural because the covered sequences as a point set is a subset of $[0 , 1]$.
This would also come out from the theorem by using $C = \{0 , 1\}$ for which C^{∞} is the set of all sequences thus covers any B and $< C > \; = 1$.
An obvious condition of when in $< B > \; \leq \; 1$ equality must stand is if B covers all sequences.
Indeed, then B covers $C = \{0 , 1\}$ and $< C > = 1$ so $< B > \; \geq \; 1$ too.
The reverse is not true, that is $< B > \; = \; 1$ doesn't mean covering all sequences. For example:

$$< \{ 0 \; , \; 10 \; , \; 110 \; , \; 1110 \; , \; 11110 \; , \; . \; . \; . \; \} > \; = \; \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \; . \; . \; . \; = \; 1$$

But it doesn't cover $111 \ldots$ and so none of the beginnings of this sequence either.

Perfect Total As Omega Number And Chance

If $< B > \le 1$ then the earlier mentioned strange limit value of $< B >$ is even more strange.
Seemingly it is now much more restricted, it starts with a decimal point without whole part.
Better to regard it in binary as an infinite sequence of 0-s and 1-s after a "decimal" point.
Then even better to regard the summing of the chances as the already mentioned **.** 0**...**01 chance binaries associated with every beginning. To add this to an already existing value means simply changing the 0 to 1 at the correct place. But if it had a 1 already then we have to change it to 0 and alter the previous 0 to 1. Or similarly go before repeatedly. So, the first 0 before will be changed to 1. This shows the potential tipping over clearly by any recognized short beginning after we approximated a beginning. So, knowing the digits tells the runnings of shorter programs.
These binary infinite "decimals" are the famous Omega numbers discovered by Chaitin.
If we are lucky and 0 or 1 are recognizable beginnings by our machine then this means a **.** 1 value added, so the first digit is established at once. Being universal implies that we can not be so lucky that both 0 and 1 would be recognized because then they could be the only collected beginnings due to B being non continuing. In reality the very short beginnings are usually not acceptable, so we can't obtain the first digit as 1. Then the accumulation is almost as bad as it was in general. So the mentioned unpredictabilities of the digits remain.
The crucial question of course is whether there are such machines at all.
To answer this, we first return to where we were.
The Kolmogorov Road was built to use measures as chances regarding sequences as points.
On a dart board throwing, the single points are the cases, here the full infinite outcome sequences.
Both of these have 0 chances but sets of these can have positive chances.
On the dart board we see the areas as measure, here with sequences we see the beginnings as obvious guide and we also assume lengths on [0 , 1].
Can we test the $< B >$ value as chance from the beginnings directly?
To make trials from the set of all beginnings seems unfair because short ones are preferred.
But the non continuingness does mean a certain fairness. Namely, if we go in increasing length then at least we don't interfere with continuing alternatives because there are no such.
So if we regard the possible beginnings as consecutive random trials of 0 or 1 that is as coin flips, then a success, that is finding one in B means the perfect one. So the chance of a termination is actually the $< B >$ chance value. We can make many such randomly generated beginning trials and we seemingly have a perfect empirical way to tell $< B >$.
But observe that failure means flipping the coin for ever and not finding a beginning in B.
So this is not a real empirical method. On the good side, this shows an amazing analogy with the machines as recognizers. They also stop or run for ever. Of course, this would be a workable coincidence only if there are machines that avoid recognizing continuations. Chaitin was one of the discoverers of these machines, for his deeper goal of showing that information lies behind.
The non continuing beginning sets, thus could also be called "self delimiting" sets because this is the term Chaitin used for such machines.
Then he used a universal self delimiting machine to collect non continuing beginnings.
This is how we started this section, emphasizing the strange total value being even stranger.
An alternative meaning of this sum value, is the above mentioned probability testing method.
Feeding randomly generated beginnings to the machine and see if it terminates.
In short, the total is also the halting probability of the universal machine.
This more dubious meaning is used as name and definition of the omega numbers by Chaitin himself. The much more important concreteness as sum is only mentioned as technical detail.
The fact that the concreteness exists in general is never even mentioned.
The already mentioned deeper false philosophy of over emphasizing information is actually covered up by this "probability meaning".

Perfect Partitioning

B^0 will be the collection of all those beginnings in B that are not continuations of others in B. In other words, these are the minimal beginnings of B. It is obviously a non continuing set. Taking these off from B, that is in $B - B^0$ there will be new minimal elements which were actually the ones that continued exactly one shorter beginning in B. So $(B - B^0)^0$ could be abbreviated as $B^{(1)}$. Then again we can take off these and the minimal elements are $B^{(2)}$. These are the beginnings that had two sub beginnings in B. And so on, we created a new: $B = B^0 \cup B^{(1)} \cup B^{(2)} \cup B^{(3)} \cup \ . \ . \ .$ partitioning. This is much better than the length partitioning was, in fact I'll call it the perfect partitioning of B.
There are six features that show this perfection.
The first trivial one is that the members are all covering each other:

$$B^{0\,\infty} \supseteq B^{(1)\,\infty} \supseteq B^{(2)\,\infty} \supseteq B^{(3)\,\infty} \supseteq \ . \ . \ .$$

The next one is a simple beauty missing from the length partitioning:
A sequence obeying B that is having infinite many beginnings from B does not mean much for the length partitioning. The sequence skips through many length groups not using them.
The perfect partitioning is used perfectly, picking exactly one beginning from each.

The third feature is that the perfect partitioning even resembles the obvious beauty of the length partitioning, the chances. Indeed, the members are all zero exponents that is minimal, that is non continuing sets. So their totals are the chances of the covered sequences by them.
Both the Kolmogorov Road and Chaitin's self delimiting halting chances try to give convincing arguments. In spite of all these "justifications", I will keep using merely $< B^0 >$ and not $| B^0|$ for the chance totals of minimal sets including $< B^{(n)} >$.

The fourth perfection of the perfect partitioning is the chance meaning of the first, that they cover each other and so: $< B^0 > \ \geq \ < B^{(1)} > \ \geq \ < B^{(2)} > \ \geq \ . \ . \ .$

The fifth perfection is a size bound for the members, following from the above non decrease.

$< B > \ = \ < B^0 > + < B^{(1)} > + < B^{(2)} > + \ . \ . \ .$ from which trivially: $< B^0 > \ < \ < B >$

$$2 < B^{(1)} > \ \leq \ < B^0 > + < B^{(1)} > \ < \ < B > \qquad \text{so:} \qquad < B^{(1)} > \ < \ \frac{< B >}{2}$$

$$3 < B^{(2)} > \ \leq \ < B^0 > + < B^{(1)} > + < B^{(2)} > \ < \ < B > \qquad \text{so:} \qquad < B^{(2)} > \ < \ \frac{< B >}{3}$$

In general: $< B^{(n)} > \ < \ \dfrac{< B >}{n+1}$. What's the big deal about this, you may ask.
The length partitioning was also diminishing, in fact we know that these are both fast diminishing. On the other hand we know that the reciprocals are slow diminishing, so this result looks straight out stupid and useless. And yet on the contrary it is something vitally new and crucial.
The big thing is that this rate of decrease is established by a fix formula regardless of B.
For the length partitioning for example, we can not predict how the groups will diminish.
We know they have to diminish fast but this is an abstract knowledge. To put it precisely:
We can not create an effective diminishing cover for the strange sequences from B using their length groups.
Martin Löf approached randomness or strangeness from this angle, that is requiring effective nil cover. At that time he didn't realize that $< B > =$ finite always guarantees the same through the perfect partitioning.
This is the sixth, most important perfection of the perfect cover but it needs two additional twists.
Observe that the length partition itself is machine creatable if B itself is machine collected.

To be precise, if a machine can recognize the b beginnings in B then an other machine can also recognize for any given b beginning and n length, that b is in B^n by simply recognizing b and its length as n. We can not predict the chance totals of these recognizable groups though.
Here with the perfect partitioning, the problem seems to be the opposite. We know the diminishing but we can not recognize the groups themselves. For a b beginning we can never be sure if some shorter sub beginning will be recognized, so we can not declare it to be in a group by merely the already recognized beginnings. Quite amazingly a trivial idea helps. Namely:
Instead of partitioning, regarding a narrowing of B. In a sense this is an opposite of partitioning: We create a sequence of beginning sets by leaving out more and more elements from B.
The starting member is B itself, then we leave out some B_1 set then B_2 and so on.
So a B_1 , B_2 , . . partitioning defines the narrowing: $B \supseteq (B - B_1) \supseteq (B - B_1 - B_2) \supseteq \ . \ . \ .$
In fact, this would be already a necessary step even for the length partitioning. Indeed, a cover by the length groups wouldn't even be meaningful without narrowing. But leaving out the longer and longer beginnings from B, the infinite occurrence from B means staying inside in these narrowing sets because only finite many was left out. Unfortunately, this doesn't solve the problem of the diminishing. We can not predict even an upper bound on the chance totals of the narrowing sets by leaving out the longer and longer length groups.
With the perfect partitioning, that is leaving out the B^0 , $B^{(1)}$, $B^{(2)}$, . . . groups one by one, everything becomes perfect. So this is the first twist, using the narrowing.
First of all, the partition members are still covers, so:

$$< B^{(n)} > \quad < \quad \frac{< B >}{n+1} \quad \text{covers} \quad B^{(n+1)} \cup B^{(n+2)} \cup B^{(n+3)} \cup \ . \ . \ .$$

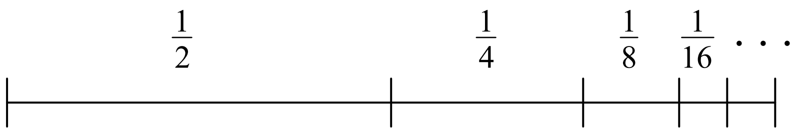
Also, to be in this (n +1)-th narrowing means that a b beginning has at least n + 1 many sub beginnings. This is recognizable by simply seeing that many already recognized sub beginnings.
So, a machine simply accepts all B elements as the start. Then when a sub beginning of anything is established it is promoted to $B - B^0$. Then if a second sub beginning is seen we can make it belong to $B - B^0 - B^{(1)}$ and so on. So, even though B^0 , $B^{(1)}$, . . . are not recognizable, these narrowings are.

But Martin Löf seemingly went much further. He required that we have narrowing beginning sets $B_1 \supseteq B_2 \supseteq$ . . each covered by non continuing, that is reliably measured C_1 , C_2, . . covers with lengths $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, . .
For a sequence, the obeying is then of course not infinite occurrence from a single B rather picking one beginning from each of the $B_1 \supseteq B_2 \supseteq$ . . sets. Or to be more visual, this means that the sequence is covered by these sets and thus also by their covers.
But why this particular halving size for the covers? Their total as 1 suggests some deep chance meaning. But it's all smoke and mirrors. Even the fast diminishing is irrelevant. These values could be the reciprocals. Indeed, from such narrowing $B_1 \supseteq B_2 \supseteq$ . . sequences we can always pick a subset that becomes such halving type and the coverings all remain.
This at once shows why the perfect narrowing is guaranteeing the Martin Löf definition too.
But we still need the second twist, namely to first avoid the < B > size itself. For this we can simply avoid finite many beginnings from B that its total becomes under 1. We can not effectively tell from the B collection when such reduced beginning set comes about but leaving out enough beginnings, such new effective collection must exist.
So this new machine or "hard" diminishing is a stronger concept than fast diminishing.
This heuristic idea was completely misrepresented in Martin Löf's original definition.

Borel's Finite Cover

Before we try to generalize the beginning concept in strangenesses, I make a detour and explain an outside generalization of the strange proof we had for $< C > \geq | b |$.
Covering the continuations of a b needs at least the size of b.
Through the Kolmogorov Road the meaning is almost trivial. The C set contains halving intervals in $[0 , 1]$ and their total length must be at least the length of a b halving interval if they cover b. But wait a minute, that was exactly our missing argument to make the visually so perfect non sequencability precise. We saw that the fractions are a sequence, we also saw that a sequence is coverable by arbitrary small interval, so an interval can't be sequencable because an interval shouldn't be coverable by smaller total. Then this "shouldn't be" was left as a later task.
So our proof solved this task but only for halving intervals. Borel showed that the same argument works not only for arbitrary intervals of the line but in space too. In fact it hides a crucial feature of our space among the abstract spaces of Topology.
Some could say, I don't need big proofs for something such obvious that a size is not coverable by smaller size. But sizes became very paradoxical, so we do need such proofs. The most famous of these paradoxes is the Banach Tarski paradox that actually doubles a ball. We'll come to this later.
Here the crucial point in our proof was that finite many intervals covered b and so for these the seeing was believing. We extended the beginnings to the longest and then still finite many covered which meant a simple counting of more halvings there, so evidently being longer.
So seemingly, the reduction of the plausibility to triviality was by going from infinite to finite.
Amazingly, the mentioned doubling of the ball also takes it apart into only finite many subsets and then puts them back to two balls. So we could doubt even our actual proof we used. But not quite!
The really last step in our proof, the doubled replacements was only possible because the pieces were intervals. So the finite replacement was crucial but became trivial through the even more crucial use of intervals as pieces not arbitrary subsets.
This doesn't mean that intervals have no paradoxical possibilities. I show you the simplest one.

We use the most trivial repeated halvings that showed: $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$:



But now we are more meticulous and tell exactly what we mean by these intervals.
Indeed, they are conflicting because the end points are shared. But not all!
The left end of the $\frac{1}{2}$ interval is its own. So we give the other end to $\frac{1}{4}$ and again and again, all will have only the left ends belonging to them. Apart from this exactification, an other simple fact is that the right end of the total shouldn't be marked on our picture because it is only approached but can never be actually covered by our intervals. Now comes the magic! I will rearrange these same intervals so that both ends will be missing. So the left end point will disappear.
In fact, I give two possible ways. First I will simply turn around each interval that is mirror to its center. So simply regard the right ends belonging to them but not the left.
Clearly the left end of $\frac{1}{2}$ will be not there now and so the total will not have that point either.
If you think that I cheated because of the infinite many mirrorings then now I will merely slide the intervals. Simply, place our $\frac{1}{2}$ interval not on the first half rather on the second. Then place the $\frac{1}{4}$ interval to its left and continuing to the left the smaller and smaller ones.

The right end remains uncovered because $\frac{1}{2}$ had no right end point. But the left end is now only approached so will be missing too.
So points can come and go, even infinite many can disappear but we claim that size remains.
We already saw an other strange end point problem with Cantor's fundamental common point axiom. Indeed, the above picture is actually the nested intervals of 1 then inside the second half then inside the fourth quarter and so on. The right end point is the only common so if we use intervals without their right ends then we have no common point at all. The solution is of course to regard closed intervals. Borel's generalization is the exact opposite, to use open intervals:

If a closed b interval is covered by a C set of open intervals, then finite many of C already covers b.

The proof is indirect and yet in a sense constructive. Namely, we prove the negative version:

If a closed b interval is not covered by any finite many elements of a C set of open intervals, then there is a P point of b not covered by C, that is by any c element of C already.

Indeed, this P will be quite constructively shown, but for the original claim this is immaterial.
The point is that b is not covered and so if we know it was then finite subset already must cover.

The finding of such P goes by Cantor's common point axiom. This sounds strange, after we just explained above that it must use closed intervals and now we use open ones. This so, because these nested closed intervals will be not from our open intervals rather we create them from the given b closed interval as starter. First we halve b and ask if the two halves are again not coverable by finite many elements of C. The crucial point is that at least one of them must inherit this property of b. Indeed, if both halves were coverable by finite subsets of C then combining these two would be still a finite system contradicting that b had no finite cover.
So choosing a half of b that is such non finitely coverable, we can repeat everything.
We halve this half again and choose a new half and so on.
These halves are regarded with their end points, so we get a narrowing sequence of closed intervals that must narrow down to a P point. We claim that it is the one we promised, that is, it is a point of b and it is not covered by C. The first is trivial because all narrowing intervals were subsets of b. Now the crucial point why P can not be covered by any c element of C.
If P were covered by a c then since c is an open interval, P must be inside c, not being an end point of c. But this means that it has a distance to the closest end of c and thus any covering interval of P that is smaller, must also be fully inside c. This contradicts that P was the common point of narrowing halvings that were not coverable by finite systems because under that size they were all covered by c itself which is a system of a single cover.

In plane the intervals can be squares and then we don't halve them rather cut in four. The point remains that one must inherit the no finite coverability. Similarly in space we can use cubes.
The theorem remains true for much wider conditions, namely for all open covers but not all closed sets to be covered. Strangely this still supports the abstractions of openness and closedness and simply specifies the conditions of real space against the abstractions of openness and closedness as the base of Set Topology. But what should open and closed mean in abstract?
An open set means that it is not bounded by a skin, it is open to the outside world if there is such.
So, every point of the set has a surrounding that is fully inside the set.
For the outside world this means that their points can not approach the set. Whatever they approach must belong to the outside itself. So amazingly, the outside is a closed set.
This complementarity of openness and closedness is a the root of the abstractions.
The other is to regard surroundings as open sets themselves and thus avoid the concept of distances too. But as I said the theorem is not true for closed sets even with distances.
Indeed, the simplest and most obvious case is if b is an infinite interval like a half line.

This is closed because every approached point belongs to it if the end point is included.
We can cover this by open intervals easily if we make sure they overlap.
But we can not cover the line with finite many of these obviously. So this infinite length ruined our claim. If a set is finite sized that is coverable by a single interval and closed too then the theorem remains. These sets are called compact. In abstract, Borel theorem can define these.

Now we still owe the explanation of a detail.
Borel's theorem beautifully showed the reduction of an assumed absurd covering of an interval by smaller cover to a finite subset but only for open covering sets. So seemingly, the whole point was missed. How can we refute an absurd covering by closed intervals? We can't just omit the end points because then the covering as condition may cease. The solution is simple.
Suppose we covered the $[0\ ,\ 1]$ interval by $1-\varepsilon$ length totaled closed intervals. Then using $\frac{\varepsilon}{2}$ by cutting it into infinite many, we can extend all used intervals at both ends and now using them as open intervals they still must cover. Also they have $1-\frac{\varepsilon}{2}$ in total.
Then we can chose a finite subset and deduce this absurdity just as well.

Alterable Sets, Boices

The reason we had to regard beginnings was because the qualitative finiteness, the machine collection can only apply to finite objects. But could we claim some plausible assumptions about the infinite sequences in general? A strangeness could be some property of the whole infinite sequences. Then of course not just the qualitative finiteness, the machine meaning is lost but the quantitative that is the chance total finiteness too. So we have nothing! But this is not true!
A new, much more heuristic principle can be formulated.
The start is to realize that the infinite sequence properties can still include claims about beginnings but if one claims something that can be concluded or rejected by a single beginning then this property can not separate random sequences from strange. To see this we need a few steps:
By the Law Of Big Numbers any possible finite segment can be produced randomly.
Letting the trials continue after this segment obtained as beginning, we at once see that any beginning is possible in random sequences.
The assumption that the continuing sequence without the beginning is again a random is an other intuition. The random sequence is random from anywhere. We can cut off any beginnings.
Adding any new beginning in front of a random sequence is a third intuition to be random.
This is a bit tricky because we can not go back in time and try to involve this beginning in a new random sequence. A more reliable cause of this intuition is if we regard the negative of being random. This is to have some strangeness that is not disappearing after we cut off a beginning by our second intuition. Now we can assume that this having infinite strangeness itself is cuttable too. Then the added beginning in front of a random sequence indeed must remain random because otherwise it would turn into non random, having an infinite strangeness, that can be cut off contradicting that the original sequence was random.
So the end result is that the complementing R set of random sequences and $\neg R$ set of sequences with infinite strangenesses are both cuttable and thus also both extendable with new beginnings.
Of course cutting and extending applied together means any alteration and so the random or non random sequences are both alterable and still remain the same.
Or to use the "alterable" for sets that are invariant for beginning alterations of the elements, both R and $\neg R$ are alterable sets.
Now comes the crucial fourth intuition that the previous "having some infinite strangeness" is actually a multitude of some concrete infinite strangenesses. This is based on the fact that we have an infinity of these as examples. For example, a sequence having all 0-s from appoint or becoming alternating 0-s and 1-s from a point. As we se adding this "from a point" to any

strangeness we get at once an infinite strangeness. This fourth intuition includes a hidden fifth that these strangenesses exclude all random sequences.
Finally we come to the sixth assumption that R is a super alterable set. Namely, for any S , $\neg S$ pair of alterable sets. One of them must be an infinite strangeness and so the other will contain all the random sequences, that is R must be a subset of it. So, we can not distinguish random sequences by any alterable collections. If a collection is alterable then it either excludes all random sequences and it is an infinite strangeness or it is the complement of such and thus contains all random sequences. This complement can also be called a naturalness, being true for all random sequences.
The only problem is to select from these beginning independent pairs which is the strangeness and which is the naturalness.
Luckily we don't have to deal with this problem because we have a much bigger one.
Namely, that the concept of collections was not defined. Using some formal language like Set Theory would be a viable road to use properties but too complex. Using Set Theory in an other sense and accept all sets of binary sequences as collections is much simpler but would contradict our sixths assumption. Indeed, lets regard a random sequence. Now lets add to it all of its beginning variants, that is beginnings changed to any new beginning. This set of sequences will be beginning independent and clearly not contain all random sequences, only the ones that end like the one we started with. But this set will not be excluding all the random sequences either because these mentioned ones are in.
This contradiction is very similar to the one we faced when we tried as Von Mises' observational sub sequences, the places after some beginnings. The beginnings could themselves be chosen by the 1 outcomes in a random sequence, thus giving back a full 1 sub sequence and defying the whole idea. Church came to the rescue with the machine idea. Now, with infinite sequence properties this is not usable, so we are at a dead end.
And yet I will go on and show that alterability is a useful heuristic concept.
First I even generalize it from binary sequence sets to any binary choice sets or in short "boices".
The two big steps are that we don't start with a sequence that is the ordered smallest infinite set rather any unordered and arbitrary large set. Then we assign to every element the 1 or 0 values.
We could imagine this as an abstract coin flip outcome in space. Or any other yes no decisions.
The first heuristic importance of this concept itself is that it can generalize Cantor's anti diagonal argument. In fact, it becomes simpler with these general assumptions. The claim is simple:

The set of boices over any S set, which is denoted as 2^S is always a bigger set than S. $S < 2^S$.

The notation itself makes sense because for a finite S set the number of possible boices is indeed 2 to the power of the number of elements of S. We used this earlier for beginnings too.
The other notation of $<$ is obvious for finite sets as their size too. That is the S elements being less than the boices in number. But what does it mean for infinite sets.

At the decimal sequences the choices were ten kind, so actually we had 10^S with S being the set of natural numbers. Using binary "decimals" the anti diagonality works just as well. In fact we don't have to worry about what to chose merely the other digit that is not in the diagonal.

So $S < 2^S$ was proved too by Cantor's method but it's still not clear what $<$ really meant.

Equivalence, Generalized Cantor Anti Digaonality

We already used isomorphism as one to one assignment between two structures that keeps all the relations. Equivalence is the purely set version without relations in the sets.
The use of this as a comparison of sizes is quite amazingly already useful for finite sets too.
If in a ball room we have hundreds of boys and girls and we want to know which of them is more, then we don't have to count them rather shout: chose! They will pair up after a few minutes of uncomfortable choices and rejections but finally which ever remains without pairs is more.
The same can be used for soldiers and horses in a field by asking the soldiers to hop on.
We might jump to the conclusion that the same idea should be used for infinite S , T sets.
That is if we can assign to all the elements of an S set different elements of T but these are only a T' real subset of T then $S \sim T'$ is an equivalence, so S is the same size as T' but some elements of T are left out from T' and so $S < T$. That seems to be the logic we used above and yet it doesn't work for infinites. Indeed, by this logic any T' subset of a T set would have to be smaller than T because T' can be regarded as S and its elements assigned to themselves.
But as we saw, subsets of infinite sets can be equivalent to the whole. In fact, as Galileo noticed even leaving out infinite many elements like all even numbers, the rest can be still equivalent to the whole set. The real logic we used above for finite sets, had a hidden twist. Namely, after we showed $S \sim T'$ we jumped to a conclusion that this makes a full $S \sim T$ equivalence impossible.
The first empirically created $S \sim T'$ merely meant that T is at least as much as S because S could be paired into T. But then we used the finiteness to jump to a logical defying of $S \sim T$.
So, the correct definition of $S < T$ for infinites is: $S \sim T'$ and $S \nsim T$.
Of course to prove an $S \nsim T$ impossibility of an equivalence between two sets is a tough thing.
Knowing that T is at least as big as S, so T will be the bigger, we assume any f function ordering T elements to S elements and we show that f can not have all T elements as value.
This is exactly what Cantor demonstrated with his anti diagonal trick. The first, empirical $S \sim T'$ meant that the naturals can be assigned into the decimals, that is the decimals are at least as many as the naturals. This was not even showed because it's almost trivial. Indeed, we can assign to every natural a decimal by for example simply putting a decimal point in front of it. This is an equivalence. It is one to one, that is unique in both directions. This is typical. Usually, the empirical $S \sim T'$ is trivial and the whole point is the refutation of $S \sim T$.
Now we are ready to prove exactly this way the grand and general $S < 2^S$.

The empirical $S \sim (2^S)'$ is again trivial by for example regarding as $(2^S)'$ the boices that are all 0 except for a single s element of S. Indeed, then to every s element of S we can assign the boice that is all 0 except for s. Now comes the real point why $S \sim 2^S$ is impossible.
Again we assume any f function that orders boices to the s elements of S and we'll show that some boice is missing as f value. So f can not be a full equivalence between S and 2^S .

Let for a b boice $b(s)$ denote the 0 or 1 value assigned to s.
$f(s)$ is a boice for every s and $f(s)(s)$ is the 0 or 1 value assigned to s by this $f(s)$ boice.
Regarding all possible s elements of S then this $f(s)(s)$ is a boice too and we can form the anti version of it too by assigning the exact opposite from 0 , 1 and denote this boice as $\neg f(s)(s)$.
We claim that this boice can not be an $f(s)$ for any s element. Suppose it were $f(t)$.
Then $f(t)(s) = \neg f(s)(s)$. But at $s = t$ it is a contradiction: $f(t)(t) = \neg f(t)(t)$.

All this beauty was a mere detour for us to show this use of the boices.
But boices hide much more.

Guessing Functions

A guessing is done by an s element of the S set to figure out b (s), that is its own 0 or 1 value in the b boice. So, obviously b (s) is not revealed to the s element but suppose all other b (t) is revealed.
An easy visualization of this is that the 0 or 1 value is visible on all individual's hat.
One can see others 0 or 1 value assignments but not one's own. Now, since the elements have no role in how the boices are done, we would think that such guessing from other's value our own is sheer luck. So it has exactly half chance to be right or wrong. A b boice without the particular value at s is the observational field of b from s abbreviated as b [s].
A g guessing function of s is giving a 0 or 1 value as guess for every b [s].
Similarly other t elements can have guessing functions and a universal guessing function G can give 0 or 1 values for every b [s] , b [t] , . . . The g guessing function of s is correct for a b boice if g (b [s]) = b (s). A universal G is correct at s and b if G (b [s]) = b (s).
It's easy to see that those b boices where g is correct and where it is wrong are two equivalent sets, so the half half chance is supported. Indeed, for every b boice where g (b[s]) is correct there is an other b' boice that is the same except at s is the opposite.
Now we can ask the question whether the different guessing functions can interrelate that is whether there is a universal G guessing function that is more successful than sheer luck or coincidence would allow. A definite sign of weirdness would be if for any b boice the different guesses for the different s were all correct. This of course is impossible because then they were correct always individually. A next best thing is that from the infinite many elements of S, the incorrect guesses are always only a finite set.
Amazingly, this is possible! The more dramatic narration is that the s elements are prisoners wearing marked hats and if they can guess their marks from the others, they avoid execution. Then assuming that they have infinite memory to have the proper universal G guessing function or their own g part of it in their heads, then they can choose. And amazingly, no matter how the prison guards selected a boice, that is mark the hats, only finite many prisoners will die.