Sets, Logic, Effectivity, Randomness

Cantor

Just as Newton's discovery of gravity needed as background dynamics, the principles of forces, Cantor's grand recognition of set sizes needed first to recognize sets as such.
The analogy continues in the sense that both new discoveries put something unquestionable on the table by interpreting major old results in a simpler way. Indeed, gravity through dynamics was able to derive the Kepler Laws and set sizes proved instantly that there have to be non algebraic or "transcendental" numbers. In fact, it showed that they are not "transcendental" at all, rather the majority of numbers. This new vision also related to a much older struggle of the Greek mathematicians to recognize and then prove that there are irrational, that is non fractional numbers. The word "irrational" again depicts something unusual though the same way they are the majority. The main vision is the same for this old result and the new.
The fractions can be sequenced, namely quite simply by the increasing totals of their numerator and denominator:

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

To sequence all algebraic numbers is a bit more complicated but can be done by the same method, that is to form longer and longer but finite groups. An algebraic number is one that is root of a polynomial equation with whole coefficients. With one T total of the order and the used absolute values of the coefficients, we only have finite many possible polynomials.
Plus every n-th order polynomial can have only maximum n many roots. So the possible roots are finite for a T too. Thus we can simply list all algebraic numbers again.
The grand contra position is the same in both problems, that is for the fractions or the algebraic numbers. Namely, that the full collection, or as we say today, the set of all real numbers, that is all infinite decimals is not sequencable.
The above almost trivial sequencing of the fractions should be a shock to a person who knows that the fractions are dense on the number line. Between any two there are others, for example the middle of the two. But the non sequencability of all decimals is more than a shock!
It is an instant infection or aversion for Set Theory. Like a first LSD trip that can make you see something you were blind to ever before but can also cause a paranoid denial of the experience.
Today the authority of history forces a person to accept Cantor's argument.
Yet in his time many did become paranoid and even wanted to silence him.
My previous argument that the new vision put something undeniable on the table was not evident for those who denied the "evidence". Later, Set Theory itself defined Mathematical Logic to be complete. So the claim that Cantor's new proof was indeed a perfect proof, seems to be a fact. But a circularity is involved. So if someone denies the mentioned completeness of Mathematical Logic in a wider sense, then he can still go against Set Theory. And Set Theory does have its own very deep problems that strangely don't affect its role in Logic.
Mathematical Logic is a revolution of the twentieth century equal to Relativity in physics.
But Logic could and should enter elementary school education and I will talk about this later.
Now I'll show the mentioned argument of Cantor to see why the decimals already between 0 and 1 can not be sequenced. So let a sequence of such decimals be:

$D_1 = .65304\ .\ .$

$D_2 = .02715\ .\ .$

$D_3 = .13902\ .\ .$

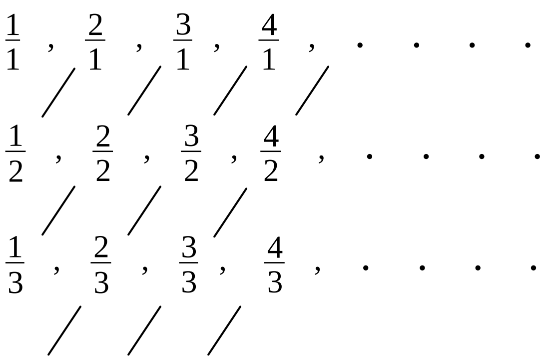
.
.

We can first create the D diagonal decimal, in our case $D = .629\ldots$
It would be strange to have this exact decimal in our list but nothing forbids this coincidence.
Now we alter every digit in D or to be even specific, we can add one to all digits meaning of course altering 9 to 0. So in our case this $D^+ = .730\ldots$
This D^+ can not be in our list!
Indeed, $D^+ \neq D_1$ because they differ in the first digit. $D^+ \neq D_2$ because the differ in the second digit. And so on, D^+ differs from all decimals in our list.
Since we were able to create a missing decimal for any list or sequence of decimals, thus the totality of all decimals between 0 and 1 is not sequencable.
Here is the part where you must stop and read the argument a few hundred times. Then the euphoria or paranoia will settle and so now you can continue.
Amazingly, these two arguments that the fractions are sequencable while all the decimals are not, is not necessary today the get a much simpler proof for irrationals than the Greeks used.
This is so because the infinite decimals themselves and the digital division process learned in elementary school are two silver platters.

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

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

Thus if someone creates an intentionally non periodic decimal, it has to be irrational.
This again suggests that the irrationals are more because there should be more non periodic decimals than periodic.
By the way, the fact that the full set of decimals is non sequencable and the rationals are sequencable, does prove the existence of irrationals but it doesn't prove that their difference set that is the irrationals is non sequencable too. This relies on the almost trivial step that two sequences together are again a sequence. Indeed, we can go through alternatingly. Same argument shows that finite many sequences together are sequencable too. Amazingly, the sequencability of the fractions actually hid the method that shows that a sequence of sequences together is also sequencable:



The sequencability itself hides the bigger concept that Cantor discovered:

Two A , B sets are same sized if there is an e equivalence between them.
Such e is simply a set of (a , b) pairs so that every a element of A and every b element of B appear in exactly one pair.

Sequencability is then an equivalence with the natural numbers.
When we listed the fractions we repeated many again and again, so to be precise we should have omitted these repetitions to give the exact equivalence.
A wider concept than equivalence is any r relation containing pairs, but the most important special relations are the functions. If an f function is defined on A, abbreviated as $D_f = A$ that is A being the domain of f, then it is already true that every a element of A will have a single b from B that (a , b) is in f. This b is what we usually denote as f (a).
The set of all taken b values, that is $\{ f(a) ; a \in A \}$ is the R_f range of f.
So an equivalence between A and B is an f function with the f (a) values being unique too and $D_f = A$, $R_f = B$.
Such f function then has a reverse or inverse function with the domain and range exchanged.
This existence of equivalence and so having same size, is abbreviated as $A \sim B$.
The used $\{ f(a) ; a \in A \}$ set collection method is quite general.
$\{ f(x) ; P(x) \}$ is the set of those f(x) assignments where x possess the P property.
This property collection was only used freely in the early naïve stage of Set Theory.
Already $\{ x ; P(x) \}$ leads to the Russell paradox and so we must restrict what properties can indeed collect sets. Besides, the f(x) assignments or functions became defined precisely too.
We avoid the full line of these examinations but we must emphasize that above the objects actually meant again sets. This is the whole point of Set Theory that everything is just sets!
This could still allow to have elemental objects that themselves have no elements.
The natural numbers 1 , 2 , 3 , . . . are visualized as such but an even more restrictive Set Theory only allows the single empty set $\varnothing$ without elements.
Then the identity or equality of any two sets is simply the identity of their elements.
An A is subset of a B in short $A \subseteq B$, if all elements of A are elements in B, that is for any s being element of A implies being element of B too. Formally: $s \in A \rightarrow s \in B$.
This implies two strange facts. Firstly that B itself is a subset which is not so problematic and we can simply call a subset “proper” if it is not itself and denote this as $A \subset B$.
More surprisingly, if A is $\varnothing$ then our formal condition is true for any B.
So, quite absurdly $\varnothing$ is subset of all sets!
The bigger picture is that we only care about collections not structure. Which means that all structures must be hidden as collections. This is not so easy!
We can not repeat same elements either. The identical elements would collapse into a single.
For example, $\{ \varnothing , \varnothing \} = \{ \varnothing \}$. Luckily this set of the single empty set is still “meaningful”.
In fact, we can again form a set by $\{\{ \varnothing \}\}$ and this would suggest to continue such set formations to define the naturals. But Set Theory had a better idea!
Indeed, a more general problem is to form pairs.
$\{ \varnothing , \{\varnothing\} \}$ is such for example but we want to form ordered pairs that instantly tell which member is the first. Here only the buildup suggests that $\varnothing$ should be the first but we may want an opposite order. The solution is to form the unordered pair {a , b} and again an unordered pair of this and something that tells the order. The simplest is to add the assumed first member, that is to define the ordered pair (a,b) as { a , {a,b}}. An other idea is to add not the first element itself rather its single set {a} , so form {{a} , {a,b}}. This Kuratowski definition is the more accepted even though the natural numbers are defined in the first manner as $0 = \varnothing$, $1 = \{0\} = \{\varnothing\}$, $2 = \{ 0 , 1 \} = \{ \varnothing , \{\varnothing\}\}$, $3 = \{ 0 , 1 , 2 \}$, . .
The above mentioned equivalence idea of Cantor could also be precisely defined only by defining functions as collections of ordered pairs.
More to the point and quite amazingly, equivalence is not just a heuristic idea to compare infinites but it is a very practical idea for comparing large finite sets too!
If for example in a ball room we want to know whether the girls or boys are more, then all we have to do is shout: “form pairs”. So then the shy lonely glances should cease and we’ll have pairs formed as far as possible. Meaning that there won’t be girls and boys standing alone.
And here the “and” was crucial because “or” in an exclusion sense can happen.

So either we have nobody alone, in the very unlikely situation that we had exactly same many girls and boys, or we have girls left if they were more, or we have boys alone if they were more.
This example at the finites suggests something:
$B \geq A$ that is B being "at least as big as" A should mean that there is an e one to one function defined on A and having values in B. Indeed, the R_e range being merely in B but not necessarily being the full B suggests the bigger or same size very intuitively.
But we also feel that if R_e happens to be a proper subset of B then we have $B > A$.
This was the very essence of the ball room situation. Yet for infinites this is totally false!
Simply because a B' proper subset of B can be equivalent to B by an f equivalence.
Thus an e equivalence between A and B can be continued into ef, picking up only B'.
Simplest example for such f is by shifting the natural numbers forward by say 5, that is ordering to every n the n+5 number. An even more "drastic" example is ordering to every n the 2n number and thus all naturals are equivalent to merely the evens. Here the left out odds are also equivalent to all naturals by 2n–1 and so actually we have two equivalent "halves".
Already Galileo observed this strange fact when he unfortunately formulated his falling distances rule by the odd numbers.
We can stretch the naturals much more to form even infinite many sequences which coincides with our already mentioned fact that sequence of sequences is still just a sequence.

The second heuristic idea of Cantor after equivalence, to compare infinites, was not to be discouraged by this stretchability of infinites, rather simply say that $B > A$ should mean that $B \geq A$ is true but $A \sim B$ is false.
This impossibility of an equivalence between A and B can be incorporated into $B \geq A$.
By claiming that any e that transforms A into B can never be an onto, that is full on B.
What's even better, we can regard any f functions for this claim instead of e equivalences because their failure means the same for one to ones and they all must fail if the one to one fail.
Indeed, combining values reduces the image.
This vision was perfectly present in our argument for the non sequencability of the decimals.
Our list of decimals was an f and we saw that it couldn't be the full set of decimals. In fact, if we regard the actual points of the number line then our f wasn't an equivalence because all 0 ending decimals after a d digit mean same points as all 9 ending ones after d – 1.
But this negative logic assumed tacitly that an impossibility of $A \sim B$ automatically implies an impossibility of $B \geq A$ too. Otherwise $A \geq B$ and $B \geq A$ both could stand without $A \sim B$.
The impossibility of this means that $A \geq B$ and $B \geq A$ implies $A \sim B$. This was not obvious to prove at all and became the famous Cantor Bernstein Equivalence Theorem.
I will not explore this because an even more subtle hidden point is more important.
Namely, that for any A , B sets, at least one of $A \geq B$ or $B \geq A$ should be true.
Indeed, otherwise such A , B sets were not comparable at all.
Cantor accepted this claim by the following logic:
We pick arbitrary a , b elements from A , B and assign these to each other as the start of our equivalence. Then we pick new elements and assign them again to each other. After any stage we pick new ones so our widening assignment is an equivalence and one of the sets must be exhausted as first. Whichever is still there is then at least as big as the one we exhausted.
The finite ball room example is interesting here because there the exhausting is not happening in an externally dictated timely fashion. The conflicts of choosing pairs was overlooked.

Zermelo

Zermelo realized that this timely vision can be replaced by purely spatial collection!
The basic $\{ x ; P(x) \}$ set collection by a P property is spatial but this is not enough!
A new spatial but not such well determined collection is to pick elements from a set of sets!
A crucial point is not emphasized here that we must pick only one element from each set.
If our sets have common elements then even if we just try to pick a single element from each, this may lead to multiples, so we can not pick a unique sample.

The simplest example is: {1,2} , {1,3} , {2,3}.
No matter how we pick "one" from each, we end up picking both from one.
Of course, if our sets are disjoint, that is have no common elements then we can pick freely and get a correct unique sample. Thus one version of Zermelo's Axiom Of Choice simply claims a choice set from such disjoint sets. Observe an important fact:
If we wouldn't require unique sample, then we wouldn't need any new axiom, simply because combining all sets will always give a weak "choice" set. Also observe that repetition of elements may not just be an obstacle to unique sample but can be an advantage to find one without the Axiom Of Choice. From {1,2} , {1,3} , {1,4} for example a perfect choice set is {1}. Something similar was emphasized by the following example:
We want to choose one from a set of different colored pair of socks.
If the socks are old fashioned being different for left and right foot, then we can simply collect the left socks in all colors as a choice set without appealing to the Axiom Of Choice. But if they are the same for left and right foot then we need the Axiom Of Choice.
But this example is misleading because in Set Theory we can not have same members in one set.
We can burry all these problems by going above with abstraction and claim not choice sets rather choice functions. Then for every set of sets we can pick freely and regard not the picked elements rather the function that orders those to the sets. Since the sets were different, the picked elements paired with each set are still different even if we picked same elements. So the pairs can not collapse. For example from {1,2} , {1,3} , {2,3} a choice function can be:
({1,2},1) , ({1,3},1) , ({2,3},2). The first two choices are the same yet the choice pairs are not.
But the problem of singular representation as such, remains hidden under these abstractions.
An amazing example how this problem resurfaced is the following:
As I said, if we allow non unique representation then the combined set is a trivial "solution".
Now we can attack this trivial solution by asking whether a non unique sample is possible that doesn't exhaust our sets rather quite oppositely, would leave unpicked elements in each set.
Then the picked sample and the left over are both weak samples, so we actually split our total set of elements so that both are such. This is exactly what Bernstein needed with elements as the points of the unit interval. Then the two halves will both have 1 covering limits and thus be non measurable. Amazingly, he still needed the Axiom Of Choice, in a non trivial manner.
So abstraction "rules" the derivabilities but I still follow visions because I believe that things will change in the future and a "didactical logic" will surface too.
Randomness of choices was buried by Zermelo's Axiom Of Choice because it seemed that this existence of choice functions is all we need anyway!
Then it turned out that the simplest question of infinites, whether there is an infinity between the sequence and the continuum is not decidable by this axiom.
So obviously the thrown out question of random choices is needed. But now this is a taboo.
The epigones just wait for a new verdict and new abstractions to follow as thought control.

Growth to Well Ordering

But I must return to the successful part of the story, how to avoid time from the above mentioned "proof" in Cantor's gradual widening equivalence.
The crucial twist is that here the use of the Axiom Of Choice is very trivial. It is the use of old fashioned property collections very smartly that does the trick of time avoidance.
In fact, a very heuristic concept can be obtained without the Axiom Of Choice.
This concept is "growth".
If we have an arbitrary f function and an arbitrary S set then the f widening of S is:
$S^f = S + f(S) = S \cup \{f(S)\}$ if f is defined on S while S itself if f is not defined on S.
Observe that this adding of f(S) as a new element could only be defined this way, by combining or union forming of the S set and the single element {f(S)} set.
If $S^f \supset S$ then we call S f widening while if $S^f = S$, then we call S f stagnating.
This can mean two things. Either that f is not defined on S or that it is but $f(S) \in S$.

Now if we start from an $\{a\}$ single elemental "starting stage" and allow f to widen $\{a\}$ as far as possible then we end up with a G first f stagnating set where $G^f = G$.
There can be other f stagnating sets but those are not obtained if we start from $\{a\}$.
The letter G reveals that this G should be our growth of $\{a\}$ by f.
But the sad fact is that this G is not describable directly because the growing stages melt into G but G has other subsets that were not stages in the growth at all.
The simple trick is to regard the "history" of G's growth, namely as the set of all stages.
So this W widening will have as its first element $\{a\}$ and as "final" element G.
Stages will then merely mean any elements of W.
The just called "final" one that we called growth, can be defined quite exactly from W.
Indeed, it has to be the widest stage, that is containing all other stages as subsets.
So combining all elements of W we must get G.
To combine all elements of W is easy by $G = \bigcup W = \{ s ;\ s \in S \in W \}$.
Indeed, this collection has all s elements that are in any S stage of W.
Before I try to give the rules for W we should visualize how really W widens:

$\{a\}$, $\{ a , f\{a\}\}$, $\{ a , f\{a\} , f\{a , f\{a\}\}\}$, . . .

These are the stages while the partial widenings are:

$\{\{a\}\}$, $\{\{a\} , \{ a , f\{a\}\}\}$, $\{\{a\} , \{ a , f\{a\}\} , \{ a , f\{a\} , f\{a , f\{a\}\}\}$, . . .

Of course, f can be "bad" not allowing these. For example if f is not defined on $\{a\}$ already or $f(\{a\}) = f\{a\} = a$ then $\{ a , f\{a\}\} = \{ a , a \} = \{a\}$ so we "terminate" here at the start.
If f is much better and is defined outside infinite many times then we obtain our first "open" partial widening as: $\{\{a\} , \{ a , f\{a\}\} , \{ a , f\{a\} , f\{a , f\{a\}\} , \ldots \}$.
Here we meant by open that this had no maximal, that is widest stage.
So here if we combine all these stages we get a new stage that didn't appear yet:

$\{ a , f\{a\} , f\{a , f\{a\}\} , f\{ a , f\{a\} , f\{ a , f\{a\}\}\} , \ldots \}$

First of all we can add this stage to the previous open partial widening to get a closed one:

$\{\{a\} , \{ a , f\{a\}\} , \{ a , f\{a\} , f\{a , f\{a\}\} , \ldots , \{ a , f\{a\} , f\{a , f\{a\}\} , \ldots \} \}$

But most importantly, from this widest stage we can continue to use f widenings.
This is the heuristic idea! To use f and union widenings until we encounter $G = \bigcup W$, the first non f widening stage.

Our proposed rules for W are:

1.) $\{a\} \in W$

2.) $S \in W \rightarrow S^f \in W$

3.) L is a beginning of W $\rightarrow \bigcup L \in W$

1.) claims the start from $\{a\}$.
2.) claims that every S stage in W has its f widening inside too.
3.) claims union widenings for all beginnings and we include W as a beginning too.

Rule 2.) does imply that $G^f = G$ because $(\bigcup W)^f \supset \bigcup W$ is contradictory.
But nothing ensures that the other stages are all f widening, that is never f stagnating.

This possibility of unwanted f stagnating stages, is only a sign of a much graver problem. Namely, that our rules only tell what we must have but they do not exclude any junk added to our intended W. So actually, our rules describe not the W widening till a first stagnating, rather all sets that contain this W. And yet, we can not give any concrete example for W.
Only the absurd set of all sets is a "concrete" example that satisfies our rules.
This gives a grand idea to get our intended W.
Namely, to apply intersection that is find the common part of all junkful of W-s.
$\bigcap Z = \{ S ; W \in Z \rightarrow S \in W \}$ collects all S stages that are stages in every element of Z.
So if Z is the set of all junk containing W-s that satisfy our three rules then $\bigcap Z$ should be the perfect junkless W. The no example problem still remains though!
So we have no clue to prove that Z is not empty, in spite of this being very evident visually.
The heuristic solution is to regard partial widenings and combine these to get $\bigcap Z$.
In fact, we'll only regard partial closed widenings, so we keep the union widening but relax rule 2.) to exempt the $\bigcup W$ maximal stage. Of course, we need some inner junk avoidance and so we not just weaken rule 2.) but strengthen it too by claiming that on all non maximal stages, f must be defined and have value outside, so real f widening can be done:

2'.) $S \in W$ and $S \neq \bigcup W \rightarrow f(S)$ is defined and $f(S) \notin S$ and $S + f(S) \in W$

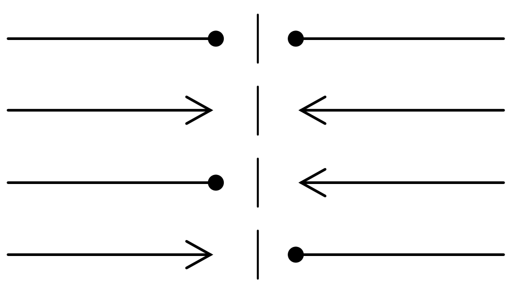
But this is still not enough and we need a rule 4.) claiming that our stages form a chain by $\subset$.

4.) $S, T \in W$ and $S \neq T \rightarrow S \subset T$ or $T \subset S$

If Y denotes the set of all W satisfying 1.) , 2'.) , 3.) , 4.) then the success of our rules should be proving $\bigcup Y = \bigcap Z$.
An immediate minor success is that our new rules have examples, namely $\{\{a\}\} \in Y$.
A trivial consequence of 4.) is that $\{a\} \subset S$ for all other S because $\{a\}$ has no proper subset.
A hidden problem we didn't mention till now, is the use of the L beginnings in rule 3.).
We might even think that time sneaked back and we are unsuccessful to avoid time after all.
Then we might think that an easy solution to define L is this:

$W[T] = \{ S ; S \in W , S \subseteq T \}$ or $W(T) = \{ S ; S \in W , S \subset T \}$.

The first is the closed beginning up to T including T. The second collects only the proper subsets of T. So T is not included and thus L will be open, that is without maximal stage, if T had no previous stage. This is the case if T was a union widening. Unfortunately, we want to claim the existence of this very T . So if we assume it in our L then we claim nothing!
It seems we failed again, but luckily we have even two options to claim union widening.
Dedekind found an ingenious way to define L without knowing what comes after it. He used it for defining continuity, but the idea is usable for any W set that has ordered elements.
We cut W into two disjoint L and R subsets so that all elements in L are left from all elements in R. In short, the L subset is left from R.
In our case where being subset is the ordering, the L are the subsets or narrower ones.
In our build up vision these are the beginning and R is the end. So:
L is a beginning of W and $W - L = R$ is the end, if : $S \in L$ and $T \in R \rightarrow S \subset T$.
Amazingly, there can only be four type of Dedekind cuts:



This picture speaks for itself if we understand that the separating "cut" line is not in the set.
The first case means that the beginning and the end have members next to each other, the dots.
So L has a maximal and R has a minimal element. This is also called as a discrete cut.
The second possibility means neither having such elements, so both approaching each other.
We could say that both L and R are limit cuts. This is also called as having a hole.
The third means L having maximal but R being limit. Here, the maximal element of L is the one that is approached by R. The fourth is reversed.
Dedekind's realization was that the continuity of points on a line simply means to have only these third or fourth kind of cuts.
So if one side approaches the other's extreme member then there is no hole!
The natural numbers ordered increasingly, will only have the first kind of discrete cuts.
The rationals or fractions can have the last three. Indeed, the cut can be at any irrational, making a hole, or at a rational in L or in R.
Our forward going widenings in W can only have the first or the last. So we must always have a next element after a beginning.
But this whole detour was irrelevant because we have a second option to avoid beginnings at all in our third rule! Namely, we can generalize our union widening for all W_o subsets of W.
Indeed, suppose W_o goes up to an L beginning. If W_o and L have a maximal S stage in them then the union of all W_o stages that is $\bigcup W_o$ is this S so it was already there in W.
If W_o and L have no maximal stage then $\bigcup W_o = \bigcup L$ because even though W_o may miss stages from L, there are wider ones that collect the missing elements. So claiming the existence for all W_o subsets will really mean the same as just for L beginnings. Thus our new rule is:

3.) $\quad W_o \subseteq W$ and $W_o \neq \emptyset \quad \rightarrow \quad \bigcup W_o \in W$

Now that our rules are exact, we can derive the two fundamental consequences from them:

I.) $\quad T \in W \;\rightarrow\; T = \{a\}$ or
$T = S + f(S)$ with $S \subset T$ or
$T = \bigcup W_o$ with $S \in W_o \rightarrow S \subset T$

II.) $\quad W_o \subseteq W$ and $W_o \neq \emptyset \quad \rightarrow \quad \bigcap W_o \in W_o$

Starting with this second, $\bigcap W_o$ is obviously the narrowest stage in W_o and the existence of such is much stronger than if we had claimed only $\in W$ as an analogue of rule 3.).
Also, an obvious consequence of II.) is 4.). Indeed, for any two S , T stages, we can regard the $W_o = \{ S , T \}$ two element subset in W and then the common part being one of them means that that one is subset of the other.
I.) seems to be a perfect junk avoidance! Indeed, we claim that all stages are either {a} or f widening or union widening. But this logic is faulty! We can have a whole infinity of junks that are all f widenings. Indeed, imagine narrowing stages that are all f widenings in reverse!
These come from "nowhere", not from our starting {a}.
This weird example explains the real role of II.) because it avoids such narrowing sequences.
But even more amazingly, we can formally "prove" that II.) is successful with I.).
Indeed, suppose we had a J set of stages not coming by our start or two widenings!
Applying II.) for this then $\bigcap J$ is the narrowest junk that has no proper junk subset.
By I.) this $\bigcap J$ has to be {a} or $S + f(S)$ or $\bigcup W_o$ with S being not in J nor the elements of W_o. Thus of course either rule 1.) was used or rule 2.) for a non junk S, or rule 3.) for the set W_o of all non junk stages. So $\bigcap J$ was not a junk contradicting $J \neq \emptyset$ at all.
This "argument" will become actual only by doing the hard yard, that is proving $\bigcup Y = \bigcap Z$.

To prove I.) lets regard W(T). If it's empty then $T = \{a\}$. If open then $T = \bigcup W(T)$.
If closed then $T = \bigcup W(T) + f(\bigcup W(T))$.
To prove II.) lets regard the L set of those S stages that are proper subsets of every $T \in W_o$.
By rule 4.) these form an L beginning and the $W - L = R$ set is the end section.
Again, depending on whether L is empty or open or closed, $\{a\}$ or $\bigcup L$ or $\bigcup L + f(\bigcup L)$ will be the minimal stage in R and W_o.

To combine the elements of Y, that is all partial widenings is easy but to claim that this the widest partial widening and so actually our widening, requires to show that this union is itself a partial widening, that is it satisfies our rules. That's why we choose such a simple rule 4.).
At the same time this rule must be strong enough to do the proof itself.
And indeed, now that we proved the II.) minimality, we can use that too.

Before we do the inheritance proof we must mention an other nuance that we ignored.
We said: "the widest partial widening and so actually our widening". What this really means is that this $\bigcup Y$ widest partial widening is then not really partial because $\bigcup\bigcup Y$, the widest stage in $\bigcup Y$ is stagnating. That is, either f is not defined on it or $f(\bigcup\bigcup Y) \in \bigcup\bigcup Y$.
Indeed, if $f(\bigcup\bigcup Y) \notin \bigcup\bigcup Y$ were, then forming $\bigcup\bigcup Y + f(\bigcup\bigcup Y)$, we would get a wider stage which added to $\bigcup Y$ would make a wider f widening than $\bigcup Y$.

To prove that all our f widenings combined as $\bigcup Y$ is an f widening, we prove a bit more.
Namely, that if we regard any X set of f widenings, then $\bigcup X + \bigcup\bigcup X$ is always an f widening. We needed this $\bigcup\bigcup X$ added stage because we might have collected widenings that are open together. For the case when X is Y, that is all f widenings, this $\bigcup\bigcup X$ is not needed because it has to be in $\bigcup X$ already. Exactly by our previous argument.

Our proof will need this claim: For any two W_1, W_2 widenings, one is a beginning of the other.
Obviously a closed beginning. And any closed beginning of a W* widening is widening.
Surprisingly, to prove this needed claim, we need a simpler version of our general inheritance that speaks only about beginnings. So it seems more general because allows open ones that are not widenings. But actually it is much more restricted because regards only an X collection of beginnings from a given W*. Then $\bigcup X$ is either the full W* or a beginning of it. Indeed:
Either all stages of W* appear in some elements of X and then trivially $\bigcup X = W^*$.
Or there are some stages in W* not appearing in X elements and the narrowest such is S.
Then $\bigcup X \supseteq W^*(S)$ because all narrower than S stages are in $\bigcup X$. But also $\bigcup X \subseteq W^*(S)$ because if a $T \supset S$ stage of W* were in $\bigcup X$ then the beginning of W* that had T as stage would have S too. So actually $\bigcup X = W^*(S)$.
Now we can prove our claim very elegantly:
We regard the common beginnings of W_1 and W_2 combined. Using our above simple combining twice with W* as W_1 and W_2, we get at once that this combined set is either the full or a beginning in each of them.
It can't be the full in both, because then $W_1 = W_2$ were. It can't be the proper beginning in both either, because then we could make a wider common beginning.
So this combined set is the full in one and a beginning in the other, which exactly means that one is beginning of the other.
Now we can prove the inheritances for $\bigcup X + \bigcup\bigcup X$.
1.) remains because every W brings $\{a\}$ into $\bigcup X$.
4.) remains because for any two S , T in $\bigcup X$ with $S \in W_S$, $T \in W_T$, one of these is beginning of the other. So the wider one contains both S , T already, where we have $S \subset T$ or $T \subset S$.
A fact continuing 4.) is that if $S \subset T$ is the case in $\bigcup X$ and $T \in W_T$ then $S \in W_T$ too.

Indeed, if $S \in W_S$ then $S \in W_S[S]$ too and $W_S[S]$ is a widening too that must be beginning of W_T because it can not be the other way around since $S \subset T$. Thus:
2.) remains because if S is not maximal in $\bigcup X$, it is not maximal in a W_T too.
3.) remains because for any $U \subseteq \bigcup X$ if a $T \in W_T$ is wider than U in $\bigcup X$ then $U \subseteq W_T$.
Or if $U \subseteq \bigcup X$ has no such T, that is goes all the way in $\bigcup X$ then the added $\bigcup \bigcup X$ makes 3.) to stand anyway.
The proof of the full W widening as $\bigcup Y$ is finished but as I said we want to show the heuristic $\bigcup Y = \bigcap Z$ claim too. $\bigcup Y \supseteq \bigcap Z$ is trivial because $\bigcup Y \in Z$.
To see $\bigcup Y \subseteq \bigcap Z$ we must show that all stages in $\bigcup Y$ must be in every $W^* \in Z$.
In $\bigcup Y$ I.) and II.) are true so if there were stages there that are not in W* then among these there were a narrowest T and it were $\{a\}$ or $S + f(S)$ or $\bigcup W_o$ with S being in W* also all S elements of W_o being in W*. But this is impossible because 1.) , 2.) , 3.) imply that then T must be in W* too.
To get equivalence between some A , B sets is now first to make W_1, W_2 full f , g widenings in them so that $\bigcup W_1 = A$, $\bigcup W_2 = B$.
To achieve this we need finally the Axiom Of Choice applied in both A , B sets or rather in their $2^A - \varnothing$ and $2^B - \varnothing$ sets of real subsets. So let c pick one element from all elements of $2^A - \varnothing$ while c' pick one from all elements of $2^B - \varnothing$.
These will not only tell our f and g but even our starting elements:

$a = c(A)$ and $b = c'(B)$. Then $f(S) = c(A - S)$ and $g(T) = c'(B - T)$.

All f values are in A , g values in B and all A subset unions remain in A and B subset unions in B.
Thus the W_1 and W_2 full f and g widenings have only stages in A and in B.
Indeed, any first outside stage would contradict that we formed them by f and g or union from “earlier” inside stages.
In particular, any non f , g widening stages of W_1 and W_2 also must be in A and B.
But here in A and B the $f(S)$ and $g(T)$ values are always outside S and T and the only sets where f , g are not defined are A and B. So $\bigcup W_1 = A$ and $\bigcup W_2 = B$.
This result is usually called as the well ordering of A and B.
The heuristic idea is then to establish a similarity between the two well orderings by assigning the starting elements and all f , g widenings and union widenings inheritingly. Then one of A or B must be similar to a beginning of the other which means an equivalence too.
The simplest technical solution to define this similarity is to regard a new third h function from f and g. Namely, h should be defined on any R set as the $(f(S) , g(T))$ ordered pair if R is an equivalence set with S and T first and second value sets. Then we can establish the W full widening again for h and it is our similarity.

Back to sets in general

This was a pretty long detour to prove the comparability of arbitrary A , B sets.
Most amazingly, this detour is also needed to prove the generalization of Cantor’s first simpler result, the sequencability of the fractions. But surprisingly, to generalize the more shocking non sequancability of the decimals is very simple and doesn’t need the Axiom Of Choice.
Both result can be formulated in abstract form by the use of set exponents, in particular with S^2 and 2^S. We already used this second as the set of S subsets but a much better approach is to regard in general the B^S set power as the set of choice functions defined on S with making choices from a B set.

So $\{0 , 1\}^S$ is then the set of all possible binary choice functions for an S set.
The 0 and 1 values could be head and tail and thus we would get the possible outcomes for coin flips at all s elements of S. A similar duality that jumps in mind is of course yes and no choices and if regard these as collection decisions for every s element, then indeed we get all possible subsets of S.
$S^{\{0,1\}}$ is the possible choices from $\{0 , 1\}$ with s elements from S as values, which actually means ordered pair formations from S. This corresponds to the fractions if S is the naturals.
The previous $\{0 , 1\}^S$ corresponds to the binary "decimals".
These meanings match the general claims that $S^2 \sim S$ while $2^S > S$.
The simpler but harder to prove $S^2 \sim S$ fact has two amazing consequences that continues the easier provable $2^S > S$. The first is quite shocking again and it says that the B base in the set power is practically irrelevant for size. This "practically" is meaningful because an obvious way to go beyond 2^S is by simply using as B base a set that is already bigger than 2^S.
Clearly if $B > 2^S$ then $B^S > 2^S$ too and this trivial way is the only to go above 2^S.
So $(2^S)^S \sim 2^S$. And this is so because quite simply: $(2^S)^S \sim 2^{(S^2)} \sim 2^S$.
The other consequence of $S^2 \sim S$ is more involved and relies on the concept of minimal well ordering for an S set. Since in a well ordered set every property has a first case, thus we also must have a first B beginning of S that $B \sim S$. If such is only S itself then it was minimally well ordered. It is then the first well ordering with that size. A longer well ordering of a well ordered S set can be regarded as a stretching of S. The possible longer and longer stretchings of S are actually all the well orderings before the next size where stretching of S can not reach. Yet together they exactly give this next size. The possible stretchings of the naturals for example are a set that can not be sequencable because otherwise we could make a new longer stretch. So we know quite exactly what is the first non sequencable infinity. It is the possible well orderings of all sequencable sets. ω is used for the infinity of the sequence or naturals while for this next infinity we use ω_1 which again can be stretched as far as possible giving ω_2 and so on. The story doesn't stop because combining the $\omega , \omega_1 , \omega_2 , \ . \ . \ .$ well orderings we then get a well ordering ω_ω that actually has to be a new size, namely a minimal well ordering of that size.
So this ω_ω came about not as next infinity by stretchings rather as limit of all earliers.
And the minimal well orderings continue after ω_ω as $\omega_{\omega+1}$, $\omega_{\omega+2}$, . .
The sad thing is that we have no clue where 2^ω, that is the decimal's infinity lies.
This 2^ω notation was a bit confusing because it was meant as set power while just above we used well ordering powers too. For 2 as base of course it is pretty obvious what we meant.
But ω^ω is really confusing. As a well ordering it is not that big, namely the limit, that is sum of $\omega , \omega^2 , \omega^3 , \ . \ . \ .$ Here ω^2 is just a sequence of sequences while ω^3 is a sequence of such sequences of sequences, and so on. So actually ω^ω is not that obvious how it looks.
But ω^ω as set power is the set of possible sequences made from the naturals. Like decimals with infinite many possible digits instead of the usual ten. By the above revealed amazing fact that $(2^S)^S \sim 2^S$ of course the in-between S^S is the same size too and so $\omega^\omega \sim 2^\omega$.
So the mystery where a minimal well ordering of 2^ω lies means the same for the ω^ω or even for the $(2^\omega)^\omega$ set powers. The only "good thing" is that we know these can not be ω_ω.
To have a better vision of well orderings, we should find alternate conditions about an α well ordering being jammed towards its end, which is obviously the case if it is a minimal.

The simplest such condition is that if we can cut any proper beginning from α so that it remains similar. ω is obviously such but $\omega + 5$ or $\omega + \omega = 2\,\omega$ is not.
Then $\omega + \omega + \omega + \ldots = \omega\,\omega$ is again uncuttable and indeed, after every uncuttable α the next is $\omega\,\alpha$. So this is merely a boring monotony again. Surprisingly, these boring uncuttable well orderings are the way to prove that $\alpha\,\alpha = \alpha^2 \sim \alpha$ which then implies $S^2 \sim S$ for any sets.
But the crucial next idea is to regard not merely beginnings that we can cut off rather any subsets that go all the way, that is not contained in any proper beginning.
These subsets are called cofinalling subsets and if an α has no smaller well ordering that is proper beginning that could be stretched out as cofinalling then α itself is called a cofinal.
This concept "almost" coincides with our previous major one, the minimal well ordering.
But not perfectly! $\alpha^2 \sim \alpha$ implies that if after α a next minimal well ordering is β then this β must be a cofinal. Indeed, if β had a smaller γ cofinalling subset then γ would be smaller in size too namely maximum α. But then the combinings of its beginnings is also maximum $\alpha^2 \sim \alpha$ contradicting that it was bigger than α.
Omitting from the condition to be a next minimal well ordering makes this claim obviously false as ω_ω shows it. So we might think that exactly just these next infinities are the cofinals.
We can prove quite easily that all cofinals are minimal but the next condition doesn't follow so it leaves open the question whether a limit minimal well ordering could be cofinal.
This would mean that it is a limit of minimal well orderings but same many must be combined as it is itself. They are called "inaccessible" and have to be so big in size that their existence can not even be proven from present Set Theory.

König

But now we still have to reveal why $2^\omega \sim \omega_\omega$ is impossible. This requires a result of König that is a generalization of the mentioned $2^S > S$ generalization of Cantor's non sequencability of the decimals. In fact, the whole idea of the B^S set as functions can be "generalized" first by allowing different B-s as the s elements of S itself. Then the functions are actually the choice functions from S and their set is denoted as $\prod S$. So we can claim that if all s elements have at least two elements, that is $s \geq 2$ then $S < \prod S$.
I used quotation mark above and indeed, this is a bad generalization. It doesn't say more rather less because this would follow from $B^S > S$ anyway and we only restricted the S set.
But this bad idea was perfectionized by König allowing to make choices not directly from the s elements of S rather some assigned $k(s)$ sets. Then with $k(s) \geq 2$ we could increase the size of not S rather R_k the range of k. But this wouldn't be anything new again.
König realized that instead of the primitive $k(s) \geq 2$ condition, an opposite $k(s) < s$ can be made and then increase the size of not R_k rather its $\bigcup R_k$ combined set.
The jump is still the $\prod S$ set but now S is the D_k domain of k, so a much nicer form of König's claim is this:

$$k(T) < T \quad \rightarrow \quad \bigcup R_k < \prod D_k$$

I even used totally new T sets not the obvious s elements of an S set because we want to show how in spite of its conditional form, König's claim implies Cantor's generalization but not directly with the s elements of S as T-s.
Namely, we should regard the k that assigns $\{s\}$ for every $T = \{(s,0),(s,1)\}$ pair made for each $s \in S$.
$k(T) < T$ because $1 < 2$ and $\bigcup R_k = S$ and also $\prod D_k \sim \{0,1\}^S$.

Indeed, every element of $\{0,1\}^S$ has $(s,0)$ or $(s,1)$ elements but replacing these by $(\{(s,0),(s,1)\},(s,0))$ or $(\{(s,0),(s,1)\},(s,1))$ we get exactly the elements for every element of $\prod D_k$.

For some $T \subseteq T_1 \subseteq T_2 \subseteq T_3 \subseteq \ldots$ sets whith $T < T_1 < T_2 < T_3 < \ldots$ we can regard $T_1, T_2, T_3, \ldots$ as domain and $T, T_1, T_2, \ldots$ as range of k.

Then $k(T) < T$ stands so we get that $\bigcup R_k = \bigcup T < \prod D_k \subseteq (\bigcup T)^\omega$.

For the $\omega < \omega_1 < \omega_2 < \ldots$ minimal well orderings $\bigcup \omega = \omega_\omega$ so $\omega_\omega < (\omega_\omega)^\omega$.

But König also proved that $(2^\omega)^\omega \sim 2^\omega$ which we already mentioned in general form as $(2^S)^S \sim 2^S$. So 2^ω can not be ω_ω.

It is interesting that König first just proved the even more special $(2^\omega)^2 \sim 2^\omega$ fact but that was already a blast. $(2^\omega)^2$ of course is merely the pairs of the decimals, that is the points of the plane and Cantor already proved that these are same many as the decimals, that is points of an interval. But his proof was very complicated. König realized that every two infinite decimal can be simply combed together with using alternatingly their digits. But quite amazingly, by using the same finite diagonal method that we showed for a sequence of sequences and lies behind the sequencability of the fractions, we can comb together a whole sequence of decimals. So the infinite dimensional cube or whole space has also only same many points as any interval. By the way, this “absoluteness” of the decimals is already present in the simplest geometrical stretching of them by a projection. And it’s quite mysterious how the individual points can become more and yet remain the same many. This relates to the ancient point paradoxes of Zenon and probably even to the mystery of where 2^ω lies if minimally well ordered.

Cantor himself thought that $2^\omega \sim \omega_1$ and this became known as the Continuum Hypothesis.
Paul Cohen proved this to be undecidable from present Set Theory. In fact, he showed that the only derivable restrictions about 2^ω are coming from König’s above theorem.
So König’s theorem is crucial and can be proven directly without Cantor’s generalization but just to see its intricacies, I will prove $B \geq 2 \rightarrow B^S > S$ first.
$B^S \geq S$ is trivial so enough to show that $B^S \sim S$ is impossible, that is if g is a function defined in S and having values in B^S then some $f^* \in B^S$ is missing as g value.
Define $f^*(s)$ as anything but $g(s)(s)$. Since B has at least two elements, this can be done.
Then $f^*(s) \neq g(s)(s)$ for every s, so $f^* \neq g(s)$ for every s too.

For König’s theorem $\prod D_k \geq \bigcup R_k$ is not trivial but very easy to show so I jump again to prove that for any $g(r)$ giving choice functions for every $r \in R_k$, there is an f^* choice function missing as g value. Let’s regard any $T \in D_k$.
Regard the restriction of $g(r)$ to $k(T)$ and collect the picked up elements in T by these $g(r)$.

This $g[T]$ subset of T can maximum have as many elements as $k(T)$.

So $g[T] \leq k(T) < T$ and so $T - g[T] \neq \emptyset$.

Make f^* by picking for every T an element from $T - g[T]$.

For any $g(r)$ regard a T that $r \in k(T)$. Then $g(r)(T) \in g[T]$ by the definition of $g[T]$.

But $f^*(T) \notin g[T]$ and so $f^* \neq g(r)$.

A strange claim that defies the Continuum Hypothesis

Now, I will show a very simple alternative condition for the falsity of the Continuum Hypothesis. But first let's regard this even simpler problem:
Suppose we have an f function that orders for every P point of the line finite many:
f (P) = { Q_1 , Q_2 , . . . , Q_n } . Prove that there have to be two P , R points so that neither is among the f values of the other. We'll actually tell how to find such P , R.
Take any P_1 , P_2 , . . . sequence of points on the line. Combine all their Q values that will be again just a Q_1 , Q_2 , . . . sequence. Since the full line is non sequencable, there is some R point outside these Q-s. f (R) is finite so there has to be some P_N not in f (R).
This P_N and R should be our pair. We already chose P_N not in f (R) so this half is true.
Also f (P_N) is a subset of the Q sequence we created and so R is outside this too.
The point was that the infinity of the sequence is less than the infinity of the full line but more than any finite subset.
In general, we have an A set and some R set of subsets that are definitely smaller than A.
If a B "between" subset in size, that is with $A > B >$ all $C \in R$ exists, then and only then :
for any f function that is defined on A and has a range inside [R] , we have two $a , b \in A$ so that $a \notin f(b)$ and $b \notin f(a)$.
And here [R] denotes the set of all A subsets that are maximum as big as the ones in R.
The easier direction is the "then" which was our example too.
Indeed, in general let's have a B and f that $f(s) < B < A$.
Let f (B) denote the combined set of the f (s) sets for all $s \in B$. This is still just B sized and so can not exhaust all elements of A and thus we can choose an a outside of f (B).
The same way, f (a) is smaller than B and so there has to be a b in B that is not in f (a).
So at once we get that $b \notin f(a)$. But also, f (b) is subset of f (B) and so a was not just outside of f (B) but also f (b) , that is $a \notin f(b)$.
For the "only then" direction we must show that if for an A set and R set of its subsets there is no B between subset, then there is an f so that there is no excluding a , b pair.
[R] now must contain all smaller sized subsets of A and this already suggests a minimal well ordering of A. This then has only smaller sized beginnings and so regarding as f (s) the beginning up to s, we have the range of f indeed in [R].
Clearly, for every s there are elements outside f (s) , namely all the elements after s.
But none of these will have an f value that excludes s because s will be in the beginnings of those.
So as simplest example, if among an infinite sequence of people everybody likes the earlier ones then everybody likes only finite many and yet there are no two that neither likes the other.
Back to the line again, if any point can like only a sequence of points then we may feel that there have to be two points that neither likes the other. But this is false because we can regard the minimal well ordering of the line and if the Continuum Hypothesis is true then the sequencable beginnings go all the way.
So with liking the members in the beginnings, we will have no excluding pair.
Of course in reverse, if we claim that there have to be such excluding pairs then this defies that the sequencable beginnings go all the way and thus the Continuum Hypothesis too.
Unfortunately, there is no plausible reason to accept this claim as an axiom though some tried to propose this.

Taboo avoidance

Now we turn to a very different application of the well ordering.
We regard something similar to choice functions that König's generalization did but sizes will not matter at all, rather the possibility that we make the choices from outside $k(s)$ sets assigned to the s elements of our S set. So $f(s) \in k(s)$ and these are the ones that we call now choice functions on S. If all $k(s)$ are the same B set then we get back B^S and our result will be just as surprising for this situation but the line of proof is just as easy for varying choices.
All $k(s)$ will be finite sets and so an immediate advantage that jumps in mind for varying choices is that this way the choices can be arbitrary big whereas with fix B it would be fix.
This advantage will be needed in a crucial application much later. Right now even fix B like the ten digits of decimals is useful. So we can visualize choice functions as "set decimals" instead of the usual decimal sequences. And this is also very important not to regard S as a given structure here at the start. What we'll claim is very spatial so we shouldn't regard any time in our choices either. A very good vision is to regard the $k(s)$ values as possible outcomes of experiments carried out at every s place. For the simplest binary $\{0 , 1\}$ outcomes we can imagine these to be head or tail for coin flips. This suggests that we will go toward some distinction of random versus artificial choice sets. But this line is also ignored just as the sizes.
So we regard all artificial outcomes as real. Quite amazingly, some things about certainty and chances will still come in. But this is only one version of the vision!
The original meaning of choices, as made by "us" will be an other and thus our results will have dual meaning. We'll even use a dual lingo for our concepts but try to use a neutral one as well.
Our first concept is such neutral, namely we'll call the finite subsets of S as windows.
If we make choices for each s from $k(s)$ but only for a w window then this is called a v valuation of w. From v itself its window is abbreviated as $[v]$.
Obviously this word "valuation" is in the vision of we making the choices and indeed an alternative name for v could be prediction if we try to guess the outcomes in a window.
The $< w >$ length of a w window is simply the number of its elements.
Observe that with binary choices the number of valuations on a w is $2^{< w >}$.
Any full f choice function or full outcome can be restricted to any T subset of S and there it is denoted as f_T. In particular the restriction to a w window is f_w.
Then $f_{[v]} = v$ means that the f full outcome regarded in only $[v]$ is exactly v.
So as prediction v was "spot on" for the f outcome. This is very unlikely, so we want to widen our prediction meaning to any V set of valuations. This V as prediction is then meant in an "or" sense, so we merely expect at least one $v \in V$ to be spot on. This of course is a clumsy meaning because our spot on meaning for v depended on any f outcome.
Yet we can heuristically define V to be "sure" if for every f outcome on S there is at least one $v \in V$ that $f_{[v]} = v$. The "failing" of a V, that is not to be sure thus means that there is g outcome on S where there is no $v \in V$ that $g_{[v]} = v$ would happen.
Amazingly, this g counterexample of V's failure could stand for g as "good", if we regard the V valuation set in an opposite meaning.
Then the v-s are not predictions by us rather given taboos to be avoided by us.
Then also the f choice functions are not outcomes, rather choices made by us to avoid all taboos of a V taboo set. And this is now meant in an "and" sense. Indeed, the g counterexample for a failing V prediction set is now a good, successful choice set made on S that avoids all taboos in V. So our fundamental first result is also dually meaningful:

Compactness Theorem Of Valuation Sets:

If a V prediction set is sure, then already some finite subset of V is sure.

If every finite subset of a V taboo set is avoidable then V is avoidable too.

First of all, observe that for a V to be sure, it is enough if V is sure on $[V] = \bigcup [v]$.
So enough to regard the $f_{[V]}$ restrictions, because the outside S elements are irrelevant.

The same goes for the second form, that is enough to find a good choice function on $[V]$.
We'll prove this second form but we'll give our g on the full S anyway because our construction goes through all elements of S. So we imagine S well ordered and will determine the choices one by one. To make choices so that taboo avoidance remains for a next beginning is not enough to ensure that this can be continued. So we need something stronger. If this indeed is stronger and implies taboo avoidance for the closed beginnings then we'll get easily that the total S will be choosable without taboos. Simply because a taboo is only finite and so it would show in a beginning with last element. So what is this stronger condition that we must require from a new choice at an s_α? That choosing any window after s_α we can evaluate it so that this added to the choices up to s_α including the new choice for s_α, we find no taboo.
This indeed implies the taboo avoidance up to s_α by simply regarding empty window after s_α.
But can such choice be made for all s_α? Suppose s_α is the first where this couldn't be done.
This means that for every choice of s_α there is some window later, that every evaluation of it added to the beginning choices will contain taboo. Let's combine these windows.
Since we have only finite many possible values for s_α this set is again finite so is a window.
Finally, lets add s_α to it too. I claim that every evaluation of this, added to the beginning before s_α will contain taboo. Indeed, if one would not contain taboo then checking what value s_α has in it and what window was used for that choice in the combining, we would contradict that all choices of that window caused taboo with this value of s_α.
This at once shows that if s_α is s_1 or a later member but with a previous member, then such failing of choice for s_α couldn't happen. For s_1 due to our condition of finite avoidability.
For having previous member because then this previous member couldn't have been chosen due to this bad window. Observe that in these two cases we didn't need that the taboos are finite.
If s_α is a limit member, we need this fact. Namely, we must choose an occurring taboo for each window choice and combine these too. This is then a finite set and so must be contained before an earlier member than s_α. But then this earlier member couldn't have been chosen.

Observe that for the first sequence of s_1, s_2, . . . members, more general non finite taboos are irrelevant because they couldn't be contained yet anyway. What's more, even the whole taboo avoidance idea can be avoided. Indeed, the condition of avoidability for finite sets, also means that every beginning can be evaluated without taboos. Regarding a set of such taboo avoiding beginnings with arbitrary length, is enough to evaluate $s_1, s_2, \ldots$ Namely:
An evaluation of $s_1, s_2, \ldots$ that "continues in" such beginning set must avoid all taboos.
And here this "continue in" means that all beginnings of the s_1, $s_2, \ldots$ evaluation is a sub beginning in a member of our set. Then indeed the claim of taboo avoidability is merely to be able to "continue in" any given beginning set if it is infinite. This was observed by König's son.
The proof is a simpler version of how to choose values. Namely, we must always choose a value so that the chosen beginning is still in infinite many member of the given beginning set.

Randomness very briefly

A stronger condition than “continue in” is to “continue by”, meaning that infinite many beginnings of the sequence must be exactly one of the beginnings from a given set.
If this set has a finite chance total and is machine generated then the sequences continuing by such set are “solovay strange”. All obvious strangenesses that we would regard as non random, are solovay strange too, with some well chosen machine, and so the negative of this, that is sequences that are not solovay strange for any machine are solovay random.
Not being solovay strange then means that the random sequences do not continue by, rather stop having beginnings from the solovay strangeness. So I also call this the “Law Of Stopping”.
This is a beautiful definition of randomness and it coincides with the three other roads that regard a sequence strange if it obeys: diminishing cover, gradual compression or successful betting sequence. Amazingly, the fifth simplest idea by Kurtz came last. This claims a sequence strange if it avoids some sure that is 1 totaled non continuing beginning sequence.
Here “avoid” means not being “covered”. And “covered” means “continue from”, that is having a beginning that is element of the beginning set.
So a random sequence can not avoid a sure beginning set, rather must continue from or occur in the set and so I call this the “Law Of Occurrence”. For example, if we claim a beginning with half chance then a different continuation with a quarter chance, then again one with an eight chance and so on, then by Kurtz’s Law Of Occurrence, one of our alternative predictions must be correct for a random sequence. We might think for a second that this should be true for all sequences but regard the following set: $\{ 0 , 10 , 110 , 1110 , 11110 , \ldots \}$.
It is non-continuing, its total chance is $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 1$ and yet the sequence $111\ldots$ will not have any beginning from our set.
Amazingly, non sequencable many sequences can be outside, that is not covered by trickier beginning sets that are still non continuing and 1 totaled.

Logic

Now we turn to the most crucial application of Taboo Avoidance.
First we need a bit of observation about the used “or” meaning in V and its negativity as falsity of their “and”. In truth, we have a deeper layer of “and” already in every v because the values in a window have to be all correct. So actually in the finite subsets of V we have a double layer of finite many “or”-s and “and”-s. This combination as fundamental came out only pretty late in Logic. So now we go back to Aristotle who made an amazing system that remained hollow. We mean Formal Logic which by the time of enlightenment was regarded as truly just a “formal” emptiness even by religious scholars.
Arguments like “If every horse is an animal and some horses are blah blah blah . . “ were indeed irrelevant for deeper thinking. So it’s quite astonishing that when mathematical logic started, the same “If-then” sentences were formed. But that was only the first stage and later by the new name of Proof Theory, the language shifted. So what stayed and what changed?
Beside this mentioned “If-then” obsession, the even more important déjà vu of the new logic with Formal Logic were the “every” and “some”, appearing also in our mockery of Aristotle.
Without getting into metaphysics, we must realize that though “some” is obviously much weaker than “every”, “some” actually hides a deeper essence of “every”. Simply because when we claim that “some . . . “ then we actually claim that this some exists. So the best is to regard “some” as “there is”. Existence or “being” and negatives as “nothing” were parts of not only old metaphysical arguments but in almost all twentieth century titles of existentialism.
It’s a tragedy that most of these thinkers never really understood the crucial new realization of Mathematical Logic and so they still talked with an ancient blindness. But what was this blindness and also the crucial failing of Aristotle in spite of his incredible foresight to regard the “every” and “there is” operations as vital? The answer in one word is, “variables”!
You may say, come on, variables were part of mathematics since ancient times!

But you are wrong! A consistent variable usage in equations and mathematical operations was not present even at the beginning of the nineteenth century. Even by Gauss!
The confusion stems from the fact that we rewrite history and quote the old mathematicians with new formalism!
Today we serve mathematics on the silver platter of this relatively new abstraction and kids already in elementary school use variables. Unfortunately, there were some idiots quite recently who tried to go against this and used blank squares, triangles and circles as variables. Thinking that this "helps the kids". But these idiots died out. (I hope!)
The fact that $x, y, z, \ldots$ can mean any number is an a priory ability of human understanding but it simply never got to this pure level earlier. It is a mystery why it took so long, because there is no biological evolution involved. An ancient Egyptian boy brought back to the present would learn math just as well in elementary school as the others. Probably would hate it as most kids do at present too, but this is a different story. It's the socially communicated language that carries the evolution! So were only these $x, y, z, \ldots$ missing from the "If every horse is an animal, . . " kind of arguments too? YES! Natural language is a maze of "cover up"-s to hide what lies underneath what we say. But what we mean and how we think, involves these!
The use of variables! We don't hide the use of "everybody" or "somebody" but we apply these quantity prefixes or so called "quantors" directly to the nouns, adjectives and verbs that hide the properties or relations of objects. So drawing attention to that "being a bicycle" or "being red" or "owning something" are merely properties or relations of objects is not enough!
We could say that B is abbreviating the bicycle feature, R is being red and O the ownership.
But applying the quantors to these directly as Aristotle did becomes hollow.
What we need is $B(x)$, $R(x)$ and $O(x,y)$. We might think that this is just stupid over complication because these variables are irrelevant. But they are not! Their choices are yes but how they interrelate with each other is not.
Plus there is a crucial application of these variable depending relations when we use them not with variables rather with fix concrete objects or persons. My bicycle is a concrete object but I don't name it though some people give names for their car. When I talk about my bicycle then I only regard myself as concrete named object as "I" hidden in "my". So my bicycle should be merely an x, for which $B(x)$ and $O(I, x) = B(x) \wedge O(I, x)$ is true.
But this naming of myself as "I" is actually a still not perfectly solved problem because everybody's sentence means a different "I"! Using Peter or P as name is much clearer!
Obviously, there are many Peters but when I talk about Peter, I usually mean a concrete person.
To say that somebody is called Peter is again a controversial claim because we can not use the Peter in place of him yet. So the name assignments and subjective "I" sentences are a complicated and excluded field in our Mathematical Logic. Luckily, most claims about the world are not such and can be perfectly described. So Mathematical Logic is not mathematical at all! It should be called Grammatics and taught in elementary school second grade!
This ain't gonna happen for still few hundred years! But let's continue in the present!
The claim that Peter owns a red bicycle is this: $\exists y\,(O(P, y) \wedge B(y) \wedge R(y))$.
The bracketed three claims combined with "and" is clear but what is this ancient hieroglyph in the front? Well, it is the "there is" quantor applied for the y variable.
And that was the whole point I was on about what Aristotle missed.
Now the quantors come alive! In fact, they kill the variable they use! So y is not a variable anymore! It was needed merely to be quantized. As we say, the full expression is closed.
The other closing was by using P = Peter. So if we had started with that open too, then we get the $(O(x, y) \wedge B(y) \wedge R(y))$ two variable expression.
This could be closed by other ways and thus make other statements about the world.
$\exists x \exists y\,(O(x, y) \wedge B(y) \wedge R(y))$ for example claims that there is somebody who owns a red bicycle. Now we're cooking, especially if we introduce the other quantor $\forall$.
$\forall x \exists y\,(O(x, y) \wedge B(y) \wedge R(y))$ means that everybody owns a red bicycle.
Clearly not true in our world and even the $\forall x \exists y\,(O(x, y) \wedge B(y))$ claim that everybody owns a bicycle is false. The $\forall x \exists y\,(O(x, y) \wedge R(y))$ claim that everybody owns something red feels to be true but actually if we allow the $O(x, y)$ ownership to mean wider objects than people in its x variable then it is also false.

A tree doesn't own anything red unless its flowers are regarded as such.
Now we can come to the newest shift I talked about from the "implication" obsession to the and-or supremacy!
The "or" is abbreviated as $\vee$ or $|$ or ∇ because we use "or" in three different sense.
The first most frequent means that we only regard it to be false if neither members are true.
In short: $f \vee f = f$ but $t \vee f$, $f \vee t$, $t \vee t = t$. The $|$ exclusion "or" on the other hand is only false if both members are true: $t \,|\, t = f$ while $t \,|\, f$, $f \,|\, t$, $f \,|\, f = t$. And finally the ∇ "either-or" is only true if exactly one member is true, so: $t \nabla t = f$, $t \nabla f$, $f \nabla t = t$, $f \nabla f = f$.
For the controversial $\rightarrow$ implication, strangely there is no controversy at all!
It is only to be false if from true assumption we get false consequence, so:
$t \rightarrow f = f$ but $t \rightarrow t$, $f \rightarrow t$, $f \rightarrow f = t$. Thus these statements are true:
If the Pope is a woman then the world is round.
If the Pope is a woman then flies can talk.
If the Pope is a man then then the world is round.
And this last shows that the truth doesn't imply any cause and effect relation.
Observe that $\_ \; ? \; \_ = \_$ can have sixteen possibility as truth definition for $?$ because for the four possible input combinations we can have this many different t , f assignments.
We only encountered five: $\wedge$, $\vee$, $|$, ∇ , $\rightarrow$ and a sixth is the frequently used $\leftrightarrow$ which suggest it's meaning, "if and only if" or "then only then" which are the same.
Clearly, $A \leftrightarrow B$ should be $A \rightarrow B \wedge B \rightarrow A$ and this raises the question of really how many such operations are needed to express the others. But we still didn't mention the most important truth operation which uses single input, the negation, abbreviated as $\neg$.
Amazingly, any single truth or false valued operation would be enough to express all others with using negations and this would suggest the implication obsession and indeed there were formalizations using only $\rightarrow$ and $\neg$. But then again we could use only $\vee$ and $\neg$ too.
The replacement of $\rightarrow$ with $\vee$ is interesting: $A \rightarrow B = \neg A \vee B$ says that the assumption is false so we can not use it or the consequence has to be true too.
$A \rightarrow B = \neg (A \wedge \neg B)$ is much better in spite of being longer because it really says what $\rightarrow$ means, being false only if a true assumption implies something false.
But the really important rules for us are these:

$\neg (A \wedge B) = \neg A \vee \neg B$ and $\neg (A \vee B) = \neg A \wedge \neg B$

$\neg \forall x \,..\, = \exists x \neg \,..$ and $\neg \exists x \,..\, = \forall x \neg \,..$

$A \wedge (B_1 \vee B_2 \vee .. \vee B_n) = (A \wedge B_1) \vee (A \wedge B_2) \vee .. \vee (A \wedge B_n)$ and

$A \vee (B_1 \wedge B_2 \wedge .. \wedge B_n) = (A \vee B_1) \wedge (A \vee B_2) \wedge .. \wedge (A \vee B_n)$

The first four allow to drive $\neg$ inward reaching finally the basic relations.
The basic relations or their negations are also called literals.
The last two allow to drive either one of $\wedge$, $\vee$ inward again and so we can end up with the already mentioned double layer of them either one being in the bottom.
Just like at the predictions, we'll prefer an "or" of "and"-s.
There is also a way by changing variables to drive the quantors outward.
These two, bringing all quantors in front and moving the $\wedge$-s in, are not essential for the system that I will show, only the moving of all $\neg$ in, to form literals.

The above rules of quantor negations forces me to tell this story of my first Analysis class at uni in Budapest. We got a new lecturer Laszlo Czach who studied in the Soviet Union.
He wrote on the blackboard the following sentence:
"Every woman has a moment in her life when she'd like to do that's not alright."
It sounds much more rhythmical in Hungarian and as Czach explained, it was from an old song.
He also said that this will be crucial to understand analysis which made everybody giggle.

Then he asked who could tell what the negation of the sentence is.
Many didn't know but soon we all agreed that it is:
"There is woman who has no moments when she'd like to do that is not alright."
Let's formalize!
W (x) = x is woman,
A (x) = x is alright to do,
L (x , y , z) = x would like to do at y time z . Thus the original statement is:

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

Let's see the negative by our rules:

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

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

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

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

There is a woman and at any time for anything if she'd like to do it, it's alright.

We of course prefer not to use $\rightarrow$ so:

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

There is a woman and at any time for anything, she wouldn't like to do that, or it is alright.

Situation matrix

The system of Logic I will reveal, uses this language but as I said we can go even further.
We can bring out all quantors to the front and all the appearing $\wedge$-s down to the literals.
These $\wedge$-s of literals could be called scenarios because their $\vee$ combining is the whole claim which we call a situation. Actually, we can avoid to use $\wedge$ and $\vee$ at all if we separate the literals in the scenarios by commas and write the scenarios under each other.
So the full claim becomes a situation matrix preceded by the quantors:

$$\forall x \ \exists y \ . \ . \ \begin{bmatrix} B_5(x, y) \ , \ \neg B_3 \\ \neg B_2(z) \ , \ B_1(x) \\ B_4(y) \end{bmatrix}$$

Now I show a classic statement of Euclid to be expressed in this manner.
It claims that there are infinite many primes.
The primes are numbers above 1 that are not composites and to be composite means to be a product with again numbers above 1.
So we'll clearly need the $>$ bigger relation and the $x \cdot y = z$ multiplication relation.
The claimed infinity can be expressed with $>$ by saying that there is prime above any number.
In fact, we can avoid the condition of primes to be bigger than 1 too because being above any value clearly implies to be above 1 too.
So we merely have to claim that there is a non composite y above any x :

$$\forall x \, \exists y \{ y > x \, \wedge \neg (\exists z_1 \exists z_2 [z_1 > 1 \wedge z_2 > 1 \wedge z_1 \cdot z_2 = y]) \} =$$

$$\forall x \, \exists y \{ y > x \, \wedge \, \forall z_1 \forall z_2 \neg [z_1 > 1 \wedge z_2 > 1 \wedge z_1 \cdot z_2 = y] \} =$$

$$\forall x \, \exists y \, \forall z_1 \forall z_2 \{ y > x \, \wedge [\neg (z_1 > 1) \vee \neg (z_2 > 1) \vee \neg (z_1 \cdot z_2 = y)] \} =$$

$$\forall x \, \exists y \, \forall z_1 \forall z_2 \{ [y > x \wedge \neg (z_1 > 1)] \vee [y > x \wedge \neg (z_2 > 1)] \vee [y > x \wedge \neg (z_1 \cdot z_2 = y)] \}$$

This is ready to be written in matrix form but can be written nicer if we realize that not being greater than 1 actually means being 1 and we use the $\neq$ symbol for $\neg (z_1 \cdot z_2 = y)$:

$$\forall x \, \exists y \, \forall z_1 \forall z_2 \{ [y > x \wedge z_1 = 1] \vee [y > x \wedge z_2 = 1] \vee [y > x \wedge z_1 \cdot z_2 \neq y)] \} =$$

$$\forall x \, \exists y \, \forall z_1 \forall z_2 \begin{bmatrix} y > x & , & z_1 = 1 \\ y > x & , & z_2 = 1 \\ y > x & , & z_1 \cdot z_2 \neq y \end{bmatrix}$$

Amazingly, every mathematical statement is such quantized matrix, using of course other basic relations than our $y > x$ and $x \cdot y = z$ were and using other names than our 1 was.
In fact, names are the key to develop a system of logical rules that guarantees truths without knowing the meanings. Truths should be valid in all realities but to call them logical necessities suggests that we need a derivation of these without checking realities.
And amazingly, the two quantor’s trivial meanings are enough to find such derivation system if we combine this with name replacements or concretizations:
$\forall x \, .. (x) \quad \Rightarrow \quad .. (n)$ for any n name that we already used or is new.
$\exists x \, .. (x) \quad \Rightarrow \quad .. (n)$ only for any new n name that we have not used yet.
The dots stand for further possible quantors. The $\Rightarrow$ symbol says that we can pick any statement on the left from our set of accepted truths and then add the right one to it too.
In the bracket abbreviating the matrix, the replacement of all x appearances by n is meant.
The crucial difference is that for $\forall$ this is totally free while for $\exists$ it is actually a new name introduction.
So we imagine a set of statements as our increasing knowledge.
Initial statements can be given as axioms and we can increase our set by these two rules.
These are the safest claims that definitely must follow from already established truths.
Indeed, the meaning of $\forall$ is that all objects obey it so the named ones too
For $\exists$ we can not assume that just because we claimed something to exist, we had name for it already but to introduce a new name for it should not cause any problem.
It should seem quite unbelievable that these safest rules are actually enough to derive all truths.
If we use our rules, write down our statements then this could go on forever, producing infinite many statements. Now suppose we observe that we have finite many $C_1 , C_2 , \ . \ . \ . \ , C_n$ among our statements that:
1. These are all totally concretized that is contain no quantors in front at all.
2. No matter how we choose a line from each of these concretized matrixes, there will be two lines that contain both $B_k (..)$ and $\neg B_k (..)$.
These dots now abbreviate the same exact names because just the negated version wouldn’t be surprising yet. But with same names of course these two can not stand for sure.
Observe that 2. claims the existence of such contradicting pairs with every possible line selections. Because, the matrixes mean “or” scenarios.
So just to find contradicting lines is not so surprising at all.

Only if all scenario selections can not escape contradiction do we have a real contradiction. Finally, observe that we said “there will be two lines that contain both . . ”.
This is a bit confusing as if we had suggested that each will contain one. And indeed, this is the typical but we allowed that already one line contains both. That of course means that that line is already contradictory in that matrix. But then how come we didn’t avoid it by choosing other?
That’s the point, maybe we couldn’t because the others would bring about other contradictions.
So the possibilities are complex but one thing is sure. This observation makes our whole set of statements useless. You could complain that our two rules that were promised to get necessities merely lead to a possibilities of contradictions.
And yet this will not only lead to our promise but even to see that our method is perfect!
So not only shall we find necessities but we’ll see that we can find them all.
The fundamental trick is to be able to make a sequence of concretizations from any set of statements so that if we check the increasing beginnings of this sequence for above described contradictions and fail, that is we always find line selections that avoid contradiction, then no other ways of concretizations can produce contradictory set for sure.
Strangely, the proof of this claim is something even deeper coming out of our sequence!
Namely, our names that we introduce in this sequence of concretized matrixes will become a reality for our full initial set of statements. Then indeed, other roads of concretizations couldn’t produce a contradictory $C_1, C_2, \ldots, C_n$ set because our concretizations are true in all realities including the sequential one that we created. So there all these $C_1, C_2, \ldots, C_n$ should be true too which means that there have to be non contradicting lines as “or” meanings.
The simple but amazing fact about negation is that if something is not true then the negated claim is definitely true. What’s more if an S statement is contradictory with other claims then in any reality where those other claims are true $\neg S$ will be true. So $\neg S$ is necessary in all those realities. So then to find out if a suspected S statement is indeed a necessity of a set of statements, we “merely” have to negate it, add this to our set of statements, start forming concretizations and hope to find a contradictory finite set.
Aside from needing a hunch to suspect such S, I had to use quotation mark for “merely” because this really feels like a wild goose chase. But using our previous special tricky sequence for our statements plus $\neg S$ we are in little better position because if our hunch was right then we must encounter a contradiction in our sequence beginnings.
Plus the underlying deeper argument shows that this method of finding necessities of a set of statements is perfect. So if we fail then no other method could show that S is a necessity.
Indeed, in the mentioned sequential reality now all our statements plus $\neg S$ are true. So if some other “logic” would claim S to be a necessity from of our statements then it would claim something to be necessary everywhere that in fact can be seen to be false somewhere.
This perfection of our Logic is called its Completeness. Meaning that it derives everything that it could or should. More concretely, if it doesn’t derive something then it is false somewhere.
A much simpler competence is of course validity, that it shouldn’t derive false or doubtful statements. So if it derives something then it has to be true in every reality.

New Geometries

When Mathematical Logic wasn’t even born yet, the concept of reality was already awakening.
Much earlier, Euclid laid down axioms for Geometry and actually this initiated the new awakening of realities. His axiom for parallelity was pretty ugly and already he thought maybe it could be derived from the other much simpler axioms. Later many tried continually and the search was exactly our indirectness. They simply added the negation of the parallelity axiom to the others and started to search for contradiction. Had they found one then they could easily reverse the arguments into a direct proof of the parallelity axiom. But instead of contradiction, strange but beautiful possibilities opened. Janos Bolyai wrote: ”From nothing I created another new world.” Gauss went on the same road and wrote a very rude reply to Janos who turned to him with his results. But as rude Gauss was, he was right. These mere possibilities meant nothing tangible yet. The coin didn’t drop neither for Gauss nor Bolyai because they envisioned their new worlds as alternate physical reality behind geometry.

Few decades later the German mathematicians realized that using not real lines as the "lines" that Euclid's axioms talk about, rather intervals or circles and so on, the parallelity axiom can be false. The most embarrassingly simple example was actually known for hundreds of years as spherical geometry. Perfectly visual on the surface of our globe. Here the lines are the main circles like the equator while smaller circles remain circles. The crucial tricky "point" is to regard the opposite points of the globe as single points. Then it becomes true that two lines cross in a point, two points determine a line, and so on. But voila, there are no parallels at all!
All main circles cross. Unfortunately, here other simpler axioms fail too. Namely, that the lines are infinite long. But there are models where distorting the lengths this fact remains formally.
And yet the coin still didn't drop after these complicated models.
Finally, it was Beltrami who realized that they had already found what they were looking for.
Observe that the origin of the contradiction searches was to see if the parallelity axiom could be derived from the others. Now if there are models where all the other axioms are true but the parallelity is not, then "game's over"! Indeed, if the parallelity axiom would follow from the others then this would mean that in all realities where the other axioms are true the parallelity is true too. But some of these inner models showed the opposite.
So for reality, mathematics doesn't have to go to physics, it has its own realities inside other realities too. Probably this new vision of reality was fermenting the concept of Sets too, but only Cantor's genius could see the writing on the wall.

Logic continued

Back to our indirect concretization method as Logic, we still didn't reveal our perfect sequence of concretizations though we can suspect that it relates to our Compactness result.
The idea is this: We want our final total concretizations obey the meanings of $\forall$ and $\exists$.
This is achieved if the partial concretizations obey them. So we go deeper and deeper into the quantors and every time we create new names for $\exists$, we go back and claim them all to satisfy the earlier $\forall$ statements. It's a back and forth game and at any point we have only finite many that definitely is not obeying all the $\forall$ meanings. Yet the full sequence will overcome this because any partial concretization only goes so far. A major fault is that we only envisioned the concretizations of a single statement. In truth we have to apply the same to all the statements in our collection. If it is a single sequence, we can go in a similar tricky back and forth way through all of them. If they are much more, than we need to be even trickier.
Now comes in the Compactness Theorem! The fully concretized matrixes are the s elements and their lines are the $k(s)$ possible choices. The taboos are any finite sets of contradicting lines, that is ones that contain contradicting literals. The taboo avoidance on finite sets is our assumption of no contradiction and the Compactness Theorem then implies a choice function, that is a full selection of lines without contradiction from all fully concretized matrixes.
And voila, the literals of these tell exactly how to define our reality because there are no contradicting ones. What's more, this reality will satisfy our quantor meanings step by step through the partial quantifications way up to the original set of statements.
Now I will show an example for a perfect sequence of quantifications for a single statement.
It could be abbreviated as $\forall x \, \exists y \, \exists z \, \forall w \; (1 , x , y , z , w)$ showing that its matrix already contains the 1 name. As new names I will continue to use $2 , 3 , \ldots$
The concretizations go by groups that either apply the earlier names in the next $\forall$ quantors or introduce new names for the next $\exists$ quantors. We have only finite many quantors but the namings go on forever. Simply because at $\forall$, before we apply the new ones we also have to apply all earliers for the new names. Both $\forall$ applications of course create new $\exists$ starting statements that need new names again.
To make the notation even simpler I will avoid the x , y , z , w variables and the matrix too.
Instead, I just write name and quantor quadruples that show which variables are replaced by the names under the original quantors.
The full concretizations will have merely names.
These form a sub sequence of our sequence of partial concretizations and these are to be used for the contradiction search.

As you will be able to see, this sub sequence will be such that:
All formed partial ∃ starting statements will have a concretization for them in this sub sequence. Plus much more surprisingly:
All formed partial ∀ starting statements will have all possible concretizations for them there.
These guarantee that then the 1 , 2 , 3 , . . . names are a reality where all the partial statements are true if all the fully concretized matrixes for our sub sequence are true.
And this then implies that our single original statement is true too in this reality.
This is the heart of the mentioned circularity on the first page. Namely, that we merely define the truth of a quantification sequence exactly by these partial quantifications to be true.
An intuitive meaning of a simple ∀x∃y∀z . . . sequence is still there if we seriously concentrate and say “For every x there is a y so that for every z so and so.” But for a longer sequence of quantors we get lost. And there is a deep problem in this. For example, we don’t realize that the universalities can always be combined to the front!
We don’t even realize this for the primitive ∀x∃y∀z . . . and only an “argument” can show us that actually it implies ∀x∀z∃y . . . : If for all x there is a y so that for this every z blah blah blah then actually every pair of x and z must have a y so that blah blah blah.
A sequential claim is merely a stronger version of the frontalized universalities.
But here the order of universalities are immaterial. We should in fact use an open expression in these universal variables. And this is not a bad idea! Directly of course all the existences combine too. But the original existences could be recovered by telling one by one for each of those variables on what universalities they depend on or not. Such alternate language of quantifications does exist and can be simpler too. And this hits exactly into our present achievement, the sequential reality we’ll show. The Löwenheim Skolem Theorem creates a reality sequence inside an already existing reality. To prove that we can be much simpler if we use the combined universalities. We merely have to gradually extend the universalities from the names used in a statement. The depending existences will come by themselves. So we don’t need the back and forth steps that we use here to create the dependences.

∀ ∃ ∃ ∀ abbreviating the ∀x ∃y ∃z ∀w (1 , x , y , z , w) statement.

1 ∃ ∃ ∀ } using the original names (1) in the old ∀ statements.

1 2 ∃ ∀
1 2 3 ∀
} making new names: 2 , 3
for the ∃ statements created in the previous group.

2 ∃ ∃ ∀
3 ∃ ∃ ∀
----
1 2 3 1
1 2 3 2
1 2 3 3
} using the new names 2 , 3 in the old ∀ statements
and all names 1 , 2 , 3 in the new ∀ statements.

2 4 ∃ ∀
3 5 ∃ ∀
2 4 6 ∀
3 5 7 ∀
} making new names: 4 , 5 , 6 , 7
for the ∃ statements created in the previous group.

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

using the new names (4 , 5 , 6 , 7) in the old ∀ statements
and all names (1, 2, 3, 4, 5, 6, 7) in the new ∀ statements.

4	8	∃	∀
5	9	∃	∀
6	10	∃	∀
7	11	∃	∀

.
.

making new names: 8 , 9 , 10 , 11 , 12 , 13 , 14 , 15
for the ∃ statements created in the previous group.

When the new Mathematical Logic realized that variables are the magic wand that rejuvenates Formal Logic, it went overboard and wanted to avoid the artificial namings as contradiction and thus truth search. In fact, it wanted to avoid the indirectness, that is the contradiction search all together. The beauty is then that from the assumed initial statements or axioms we go step by step to derive new expressions. The ugliness is that I had to say expressions here! Indeed, due to the avoidance of artificial logical names, we must derive not only statements but open expressions too, corresponding to the partial concretizations. We will need logical axioms too!
Namely, all finite $\vee$ expressions that contain contradicting literals. This corresponds to matrix lines combined containing contradicting literals but now it can contain anything else too.
There it was contradiction because we had $\wedge$-s but now with $\vee$-s this means a trivial necessity. Indeed, an "or" that contains opposite members must definitely be true. The big extra baggage is all the other expressions included. But this extra baggage is exactly the secret treasure that will make our final statement that we derive. To give a nicer feel, we could say that only two opposite literals with a single $\vee$ are to be regarded as axioms and then allow an other logical axiom to widen these with an arbitrary expression. In fact, such widening should be allowed from any derived expression, so this is a more generally useful rule.
The $\exists$ introduction rule is in perfect harmony with the $\vee$ widening! Totally unrestricted!
As total opposites, the $\forall$ and $\wedge$ introductions are very restricted.
$\forall$ can only be introduced if the variable doesn't appear anywhere else.
Before I come to $\wedge$, I tell the simplest rule, appropriately called simplification.
We start from $\vee$-s in our logical axioms, so we merely need simplifying the built ones if we reach an $\vee$ combination with more same members. Obviously, we keep only one of them.
The $\wedge$ introduction is a twist on this simplification rule.
Namely, if we derive two almost identical $\vee$ combinations except of one member, A_i being B in one, so look like $A_1 \vee A_2 \vee .. B .. \vee A_n$ and A_i being C in the other, so look like $A_1 \vee A_2 \vee .. C .. \vee A_n$, then we can derive $A_1 \vee A_2 \vee .. (B \wedge C) .. \vee A_n$.
This corresponds in concretizations to simply continue separately for two statements connected by $\wedge$. We didn't see this in our matrix method because all $\wedge$-s were at the bottom.
So this system allows much more natural expressions than our quantized matrixes.
Namely $\wedge$, $\vee$, $\forall$, $\exists$ used freely and only assuming the $\neg$ negations at the bottom.
The $\wedge$ introductions mean that we don't just have to start from an ad hoc $\vee$ combination as logical axiom, but actually a whole set of such that then will merge by these $\wedge$ introductions.
Now comes the "piece de resistance"! Our ingenious concretization sequence can be incorporated by starting from any statement suspected to be derivable and obtain all the needed ad hoc logical axioms. Namely:

We'll regard the obviously unimportant order of the $\vee$-s in our suspected statement as very important! And so:
We apply variable replacements corresponding to the reversed quantor introductions always at the first appearing quantor in our $\vee$ expressions.
At $\forall$ we simply use a next new variable.
At $\exists$ we replace the "next" yet not applied variable and add the used $\exists$ sub expression with an $\vee$ to the end. This explains the used "next" word more precisely. Namely:
These added $\exists$ expressions gradually become first infinite many times. For their replacements we must use the first from our used variables that hasn't been used for this $\exists$ expression before.
Encountering an $\wedge$ we must apply the whole process for both members.

As an example I will regard the $\forall x \exists y \forall z [B(x,y) \vee \neg B(x,z)]$ logical necessity.
This contains no $\wedge$ so we have a single process.
I will use $1, 2, 3, ...$ as new variables and before I do the replacements I will prepare this by putting the new variable after the old original but still keeping the quantification.
This way the process will become perfectly clear:

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

$\forall x1 \exists y \forall z [B(x,y) \vee \neg B(x,z)]$

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

$\exists y1 \forall z [B(1,y) \vee \neg B(1,z)]$

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

$\forall z2 [B(1,1) \vee \neg B(1,z)] \vee \exists y \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee \exists y2 \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee \forall z3 [B(1,2) \vee \neg B(1,z)] \vee \exists y \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B(1,3) \vee \exists y3 \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B(1,3) \vee \forall z4 [B(1,3) \vee \neg B(1,z)] \vee$
$\exists y \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B(1,3) \vee B(1,3) \vee \neg B(1,4) \vee$
$\exists y4 \forall z [B(1,y) \vee \neg B(1,z)]$

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

$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B(1,3) \vee B(1,3) \vee \neg B(1,4) \vee$
$\forall z5 [B(1,4) \vee \neg B(1,z)] \vee \exists y \forall z [B(1,y) \vee \neg B(1,z)]$

$B(1,1) \vee \neg B(1,2) \vee B(1,2) \vee \neg B(1,3) \vee B(1,3) \vee \neg B(1,4) \vee B(1,4) \vee$
$\neg B(1,5) \vee \exists y \forall z [B(1,y) \vee \neg B(1,z)]$

This goes on infinitely!
But at the eleventh line we could have stopped because $\neg B(1,2) \vee B(1,2)$ appeared which makes that line a logical axiom. From this line we can derive our statement strictly, using quantor introductions, corresponding to the variable replacements and simplifications corresponding to the used end repetition trick:

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

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

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

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

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

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

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

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

The beauty of this system is of course the seemingly crazy thing what we did above, that is to continue the process in spite of founding a truth! Indeed, what if we had not found such truth? The reversals in negative show that then we found something even better! A sequential reality where the opposite of our statement is true. So then we can be sure that it wasn't a wild goose chase, the statement is definitely not a necessity!
To use axioms to derive something, the axioms have to be broken down to open expressions and those used. The break down rules are easy and the usage, that is mixing with the logical axioms needs only one rule. Not surprisingly, now this is a twist on the $\wedge$ introduction rule.
Namely, suppose that the mentioned B and C only differing members of two derived otherwise same $\vee$ expressions are such that $C = \neg B$. Then $B \wedge C = f$ false and so can simply be omitted from the combined $A_1 \vee A_2 \vee .. (B \wedge C) .. \vee A_n$.
So we can say instead of the $\wedge$ introduction rule, that deriving both
$A_1 \vee A_2 \vee .. B .. \vee A_n$ and $A_1 \vee A_2 \vee .. \neg B .. \vee A_n$ allows one to derive
$A_1 \vee A_2 \vee .. \vee A_{i-1} \vee A_{i+1} \vee .. \vee A_n$. So we can "cut out" the opposite members.
Observe that this "cut" rule could not be simplified by some f rule for falsity. $B \wedge C = f$ is not a derived statement, in fact if it were, we reached a contradiction. The formal opposition of the parts is only that allows the cut. What's more, in our system this formal opposition is not even formal because we have no $\neg$-s at expression levels. So our cut rule must involve a verification first that the two expressions are opposite. This of course is not that hard.
This cut rule should be called Pair Cut Rule referring to the cut out contradicting pair.
This is useful because other "cuts" are possible too.
The weaker Single Cut derives a C expression from the derived A and $\neg A \vee C$ ones.
Its meaning comes out by realizing that this second is $A \rightarrow C$.
So this is the good old chain rule or Modus Ponens.
An alternate cut that cuts out more, yet is equivalent with the Pair Cut is the Pair and Or Cut:
Here we derive again C but now from three derived expressions: $\neg A \vee C$, $\neg B \vee C$, $A \vee B$.
Again, the better meaning comes out by realizing the first two as $A \rightarrow C$, $B \rightarrow C$.
And this suggests at once a big jump to cut out many expressions if $A_1 \rightarrow C, \ . \ . \ . \ , A_n \rightarrow C$ are all derived together with $A_1 \vee \ .. \vee A_n$. Indeed, these should allow to derive C.
This Multi Cut is the general way to derive something from things that we don't necessarily derive individually. Especially because these are expressions not statements. But even for statements we can use this trick to avoid the complicated details. An amazing example is to derive that there are irrational numbers so that their exponentiation gives a rational.
The Pair and Or Cut is enough here. The first A statement should say that $\sqrt{2}$ used as both base and exponent is a good example, that is they are irrationals and $\sqrt{2}^{\sqrt{2}}$ is irrational.

B should say that $\sqrt{2}^{\sqrt{2}}$ used as base and $\sqrt{2}$ as exponent is a good as example, that is they are irrational and $(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}}$ is irrational.
C is the claim that there are irrational powers being rational.
$A \to C$ because $\sqrt{2}$ is indeed irrational so this half is true of A and if the other half is true too then we got an example for the claim.
$B \to C$ because $(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\sqrt{2}} = \sqrt{2}^{2} = 2$ so now this second half of B is true.
So if the first half that $\sqrt{2}^{\sqrt{2}}$ is irrational is true too then we get a verification for C again.
Observe that the falsities of A , B are irrelevant because these make the implications still true.
What we do need, is to see that $A \vee B$ is true.
And indeed, if A is false that is $\sqrt{2}^{\sqrt{2}}$ is rational then B must be true.
So we obtained C without knowing whether $\sqrt{2}^{\sqrt{2}}$ is irrational, which is probably true.

Already the Single Cut or Modus Ponens allows to use multiple conditions derived.
Indeed, from $B_1 \wedge B_2 \wedge .. \wedge B_n$ and $(B_1 \wedge B_2 \wedge .. \wedge B_n) \to C$ derived we get C.
This second of course is the same as $\neg(B_1 \wedge B_2 \wedge .. \wedge B_n) \vee C = \neg B_1 \vee .. \vee \neg B_n \vee C$.
So repeated single cuts can be used. Especially interesting is if these $B_1, B_2, .. B_n$ conditions are merely some members from a set of $A_1, A_2, \ ...$ axioms.
Then if $\neg B_1 \vee .. \vee \neg B_n \vee C$ is a logical necessity we can use our axioms with Single Cut.
And indeed, single cut is enough because by Compactness, finite many axioms must imply any C that is true in all realities of $A_1, A_2, \ ...$
Still the Pair Cut is a better choice than the Single Cut which needed the $\wedge$ formation.
The reason for this lied in the true motivation of Gentzen to create his new system to replace the already accepted one created by Hilbert. Gentzen's system was much more complicated than what we introduced because he didn't restrict the expressions to have all $\neg$-s at the bottom.
Ours was discovered by Tait as a simplification of Gentzen's.
The heuristic importance of Tait's system is that in spite of requiring the negations at the bottom, with allowing both quantors, $\wedge$ and $\vee$ in any order, the natural statements of mathematics or even every day sentences are very naturally translatable. So having the seen algorithm that gives a definite search for logical necessities is a crowning of this system.
But observe that the crucial point here is the "definite". Namely, that a failure of the search directly gives a model for the non necessity, that is for a falsity of the suspect statement.
Indeed, merely a search for necessity can be obtained from any system of Logic!

Generation

The even wider fact is that any derivation system is also a generation system!
The simple reason is that there are only finite many rules. A hidden problem is that we need axioms too and usually these are infinite many. But these are generable too as a sequence.
So going up to a beginning in this list and using the finite many rules for only a restricted many times we create only a finite number of derivations. Now if we increase the beginnings, that is the used axioms and the number of rule applications simultaneously in any manner, we will get all possible derivations. This is an ugly but effective sequence of all derivable theorems that mathematicians haven't even discovered yet. This feels paradoxical but it is perfectly natural with a new Effectivity vision. This vision is best understood as an opposite of Hilbert's vision about derivabilities. He dreamed about a system that would decide if something is theorem.
Observe that if the previous ugly listing of theorems were increasing in length of the theorems then Hilbert's dream came true at once.
Indeed, as we surpass a length in our list, we can find all the missing ones with same length.

And so we get the non-theorems too increasingly. Thus a decision for a statement whether it’s theorem or not is simply checking the two lists up to the length of the statement.
But our ugly list will definitely not be increasing in length! Short theorems may come much later then long ones. This comes about naturally due to the use of axioms!
Indeed, the logical rules would keep some length increase. In our system for example, all the rules increase length except the cuts. But the axioms which are just a second list, when mix with the logical rules, make the possible shortening of the derived theorems unpredictable.
The most amazing thing is that even without non logical axioms, that is just regarding the logical necessities, the situation is the same if the language has at least one basic relation.
Only having only basic properties allows a decision method for necessities.

Gentzen

Returning from this little detour to the Tait system is very educational.
We didn’t merely derive the necessities but had the reverse method to search for the necessary logical axioms too. By what we explained, the lengths of the found axioms can not be predictable by the length of the initial statement that turns out to be a necessity.
Indeed, otherwise we could tell in finite steps if a statement is not necessity by using only axioms under certain lengths and use all the increasing rules and no cuts.
So the avoidability of cuts was very useful here. This relates to what above we hinted as the “true motivation of Gentzen” to create his system. We might think that this motivation was the same that we called as crowning result of Tait’s system, the direct verification of necessity.
Except that Gentzen didn’t have to use all negations at the bottom.
But this wasn’t Gentzen’s motivation at all, “merely” the elimination of cut rules.
We had to use quotation mark because more is behind and it is in the used word “elimination”.
The crowning result proves that cut rules are not needed, but this is not an elimination as such!
Let’s start from here:
The fact that for logical necessities we don’t need cut rules, suggests a dilemma. What if we derive some logical necessities and then use those to derive a new one by using cut rules.
This obviously will be shorter than deriving the new necessity from the bottom without cuts.
So is there an algorithm that gives the long and tedious road from the simpler one?
This is a true elimination of the cuts and this was Gentzen’s goal.
To find such cut elimination algorithm, the best is to use the Pair Cut rule which is more powerful than Single Cut but much simpler than Pair and Or Cut or Multi Cut would be.
That’s why the Pair Cut became simply called as the Cut rule.
But this idea, to be able to find a cut elimination process hid an even bigger goal, beyond necessities, that is having non logical axioms, namely the axioms of Number Theory.
Derivation methods that are complete do not imply that the axioms have a reality.
Indeed, that would be absurd because we can start with intentionally contradictory axioms.
Remember, we also said that the completeness of our logic shows that statements true in all realities are derivable. But we didn’t mention if there are such realities or not.
Our sequential reality construction too was relative depending on the beginnings avoiding contradictions. Without the cut rule it is very evident from the rules themselves that a contradiction can not be derived and indeed we only derive necessities true in all realities.
So if a similar cut elimination could be found for Number Theory then this would show that Number Theory can not lead to contradiction.
You may say that this is an even crazier goal than to eliminate cut itself!
Indeed, the natural numbers as reality surely satisfy the axioms quite visibly.
Also, a contradiction would mean that all statements are derivable. Which feels absurd if we think about those hard to prove complicated theorems of Number Theory?
Unfortunately both arguments have holes.
The set of naturals obeying the axioms is only visible from the outside.
A strict Number Theory should only talk about numbers, that is finite sets.
For the complicated proofs we can reply that maybe contradictions are even harder to prove.

The fact that we have no contradiction in Number Theory is a quite complicated claim.
Namely, that there can not be derivation of both A and $\neg A$ for any A statement.
Of course, if this were the case then we could derive any S statement by widening A to $A \vee S$ also $\neg A$ to $\neg A \vee S$ and then cut to S. So formally we must show that not every S is derivable which seems so easy. But to see this, would mean to analyze all derivations.
We might even use number theoretical concepts along the line and in fact use Number Theory to prove its own non contradictoriness or “consistency”.
Gödel proved that this is impossible! First of all, a system has to be complex enough to be able to talk about derivations translated into its own language. Number Theory is complex enough and to show this was Gödel’s first and longer achievement. But then a quite simple argument showed that if this non contradictoriness is proved then it actually means a contradiction!
So if you believe in that Number Theory is consistent, then you also must accept that this can not be proven. Consistency is a question of “faith”.
But Gentzen wanted to challenge this bleak picture and see how far one can see the consistency of the axioms of the naturals. And he succeeded! Without “breaking the law”, that is contradicting Gödel’s result. What’s more, his system became the start of Proof Theory.
He was a strange person too! A Nazi who corresponded with Jews and died in a Russian labor camp. Gentzen also did something else, what a less famous person Jaskowski did at the same.
They introduced new, “more natural” deduction systems already for quantorless expressions.
The indirect concretization I stated with, was only conquering the quantor meanings.
The way I explained, the quantorless expressions are just stupid combination of the possible sixteen operations. So every such expression can be evaluated by the possible t, f combinations and seen to be always t or not. End of story!
I maintain this position but I also realize that such evaluation is actually a tremendous task.
To simplify this process is a nice goal and as computability became a subject, this side gave a real importance of these “natural” systems. But some are championing them as philosophically important subject. I just regard it as new version of the bad old Formal Logic.
Thinking, whether mathematical or other, is much too complicated to be formalized!
In my opinion it is not even material! I was passionately telling that Grammatics, the introduction of the true hidden nature of statements should be elementary school subject!
But if you propose a deduction system that reflects thinking, I just say “you are crazy”.
So how can I regard the forms of all everyday claims to be simple and yet regard the thinking about them as transcendental?
Because we don’t think through these claims. What we say is not how we think!
The claims are finalized exact forms of verifications and we like to “publish” only these.
We think in visions and they should be transferred too as well as possible!
This article is an example of trying to transfer full visions. Rampant Formalism of abstractions is the present main stream tendency not only in mathematics but gradually in all avenues of communications. But we must come to the most important aspect of derivations.

Turing

Gentzen was not the only one who tried to defy Gödel! So now we come to the historical background of the little detour we made above about the listing of all theorems.
A young English mathematician Turing, who later became the mind behind breaking the German code system in the War, realized that Gödel's vision was narrow.
Gödel's above mentioned result, the non provability of a system's own non contradictoriness was only a culmination of his first similar result. This showed that the ability of Number Theory to talk about its own derivations causes the system to have undecidable S statements for which neither S nor $\neg S$ can be derived. The truth is very different:
The ability to talk about itself indeed causes the system to be complex as derivation method.
But then this complexity plus the unemphasized fact that the infinite many induction axioms are given by a simple scheme and thus effectively, is the real cause of undecidability.
In fact, it is not the undecidability that is basic at all! Rather that the derivable statements, the theorems as a set and the underivable statements, the non theorems are different kind of sets!
That's why the effective listability or generability of the axioms is crucial.
Continuing with derivations from them in some fix order that goes through all possible derivations, these two, the simple generation of the axioms and the complex generation of all possible theorems, form a single complex generation of all theorems.
And amazingly, then the complement set, the non theorems is an even more complex set that can not be generated by any method at all!
This is what forces the existence of undecidable statements!
Indeed, if there were no such undecidable statements, then we would have an immediate method of generating all non theorems by simply generating the theorems and then just formally negate them. I remember when I realized this whole vision from Rosser's book after reading all the junk about Gödel before.
Of course, there is a vital point here that we must reveal and makes sense of why this grand vision is not emphasized, in fact is covered up. This reality of the non theorems of Number Theory being a "non generable set by any means" is not a mathematically provable claim.
Simply because the "by any means" is not a definite claim.
Effectivity is a mystery, still outside mathematics!
Gödel's proof on the contrary was totally exact. It didn't mention Effectivity, generations and such semi physical concepts. It merely looked at the derivabilities of Number Theory, coded by numbers. Number Theory then can talk about its own derivations and this implies the undecidable statements.
For one who dares to talk about the reality that the non theorems are a non generable set, this means that he not only hid the fact that the set of non theorems is non generable but that those related number sets are such too.
Turing went to the bigger, semi physical vision and yet he also stayed within mathematics!
By replacing Number Theory with something that at least suggested the wider field of Effectivity, namely with Computers. In fact, he used the "computable numbers" expression already in his title. We might think that he was merely sloppy to say computable sets of numbers meaning the generable code numbers that have non computable complements.
But this was not the case at all! He meant real numbers in the title, which are infinite binaries and so correspond to set separations of the naturals. By computable then he meant both half being generable which was in compliance with the then accepted vision of Recursive sets also meaning both half calculated. This choice for the word computable, preferring real numbers and not sets of naturals has a very nice consequence though. Namely, we can return to the very start of our article, how Cantor generalized the existence of irrational numbers to transcendental.
Fractions are special algebraic numbers and algebraic numbers are special computable numbers.
So Turing went even further than Cantor by showing that there are non computable reals.
Most amazingly, even the method is similar, an anti diagonal argument.
The big difference between Cantor's and Turing's anti diagonality is that this second doesn't imply a bigger set. So the sheer existence of non computable reals begs the question of whether these non computable reals are a majority in some sense.

Randomness again less briefly

And indeed, our heuristic infinite decimal vision suggests that in fact, a majority of the reals should be actually random! And observe that this intuition regards random to be much stronger than merely not computable, namely lacking any predictability or strangeness at all.
This new meaning of majority versus minority was discovered by Kolmogorov already in his Probability Theory that actually wanted to avoid Randomness.
His student Martin Löf realized that we can talk about random versus strange sequences by exactly using this cover distinction combined with effectivity.
The minority, the strange sequences are more basic than the majority, the random ones.
For two reasons. Already this Kolmogorov duality is simpler for the minority as being covered by a diminishing beginning set. Usually such minority is called as a nil set.
Being covered by a beginning set means to "continue from" the set, that is having a beginning there. This means not much.
Yet as I already mentioned in the previous brief section, the latest Kurtz strangeness only relies on this. Namely, by a sequence disobeying the Law Of Occurrence that is avoiding an effective sure non continuing beginning set.
But in the Kolmogorov nil sets the sequences are covered by a whole sequence of beginning sets which in addition have to have diminishing chance totals.
The other reason why strangeness is simpler, is that it can be grasped individually as something effective that a sequence obeys, though in Kurtz's case it is more a disobey.
In Martin Löf's case the obeying is being covered by diminishing beginning sets so that the sequence of these beginning sets is effective and also their diminishing is effective.
So in short, a Martin Löf strangeness is an effectively nil set.
If we only effectivize the set but leave diminishing as a normal, then we can call the covered set an effective nil set which is thus a wider strangeness and so leads to a narrower randomness.
This already suggests that the seemingly perfect three identical alternative definitions are a false over simplification. The real beauty is then that this splintering of these strangeness and randomness concepts are intimately interrelating. For example, this wider strangeness of effective nil sets relates to the compressibility splintering. I will explain now how:
Chaitin the biggest champion of regarding incompressibility as "the only true definition of randomness" himself caused the doubts in his vision by discovering the famous omega numbers. These are determined by a machine and form effective nil sets as single sets.
Only not diminishing effectively is that avoids them to be strange by Martin Löf's definition.
But many regard them to be strange and accept the effective nil sets as strange in general.
If your head is spinning it's understandable. How can something determined by a machine even be contemplated as random? Well because the word "determined" is actually undefined yet.
I used it merely as a short for what I will explain, namely how it is determined.
If I choose ten random numbers and assign them to ten machines one by one, then each machine will determine one of the random numbers. This of course is not determined by the operations of the machines. The omega numbers are, but to make things even clearer I will start with a hypothetical intermediate machine determination. Not yet relating to the omega numbers neither to diminishing beginning sets. Then I will make the connection by showing machines toward both. One that will use our hypothetical, to generate beginnings so can be used to define effective nil sets, and one that can produce our hypothetical from beginning generators.
Our hypothetical machine will use every n natural as input and will generate 0-s and 1-s.
For better visualization, imagine the naturals as a heading and under each, the generated 0 , 1 digits as columns. Now the big assumption is that we know each will fixate from a point.
So our machine will generate only 0-s or only 1-s in every column under a point.
As you could guess, these fixated digits will be the digits we use as our real number.
You can not argue with the perfect determination, only whether such machine exists at all.
But if so, then should this number we created be random or not?
Most would say not, and so defy Chaitin's faith in his vision.
Now first comes the second part, to use our column machine to collect beginnings because all strangeness definitions rely on this and not on the individual digits. So we do the following:

For every n value we let our machine go for all the 1 , 2 , . . , n inputs exactly n steps. We ignore this initial running. Instead, we collect the n long beginnings that show up after, that is under. Infinite many are there but after finite more steps than n, all combinations must come about. If none of the 0 , 1 values fixated in the n steps then we get all possible n long binary beginnings. But if some fixated then those will only appear with those digits.

The chance value of an n long beginning is $\frac{1}{2^n}$ and so if all 2^n many beginnings appear then the chance total is 1. But if one digit is fixed then the total is only $\frac{1}{2}$ because we have half of the cases. With more and more digits fixed the total chance is diminishing. Gradually, all the beginning digits are fixed too, so our beginning sets will cover only our fixated digits.
So our sequence as a single set is an effective nil set. But not an effectively nil set.
Because the diminishing was not effective. We had no bounds by n how small the chance totals are for the n group. We merely deducted the diminishing from the fixations.
It's interesting to see if the alternative solovay strangeness fails too as it should. Infinite many of the groups can be combined into a single and our sequence must continue by this. This is okay but the total chance must be finite too. And this we can not guarantee without knowing how the group chances diminish. For example: $\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ . \ . \ . \ . \ . = $ infinite.
Now we will show that our hypothetical fixating machine can exist.
Namely, can be manufactured from a simple beginning generator that generates a set with finite chance total. And such must exist because generating only finite many beginnings is such.
So we have such beginning generator and then merely attach a simple calculator that adds up the chances as the beginnings are generated. Finally we display this total in binary form.
Voila, the value of the total will fixate as more and more beginnings are generated.
In fact, the digits must fixate in such way that if one is fixated then all left ones are too.
A minor problem is that we don't know how big is the whole part of our infinite binary and so for a while the positions of the digits is jumping too. But once the whole part or at least the first 1 digit of this is fixated, we will not shift forward.
A more readable binary total could be obtained if we knew that our total is definitely under 1.
If our machine generates beginnings that are never beginnings or continuations of an earlier generated one, then this stands. And so, such can be made from any machine by simply omitting newly generated beginnings that are beginnings or continuations of earliers. For this of course we must store all earlier generations. Which is okay because we have infinite memory!
This randomness direction was a drastic continuation of Turing's non computability, that is the non generable complements. Indeed, random reals, that is infinite binaries, are separations of the naturals by the 0 and 1 digits so that obviously both halves must be non effective.

Rice very briefly but very strongly

The other quite opposite direction, where we know one half to be effective but realize that the other half is usually non effective was that we regarded as the cause of Gödel's, undecidable statements. The most important result of mathematics after the war was in this direction and is called Rice's Theorem. I regard it as an equal of the Well Ordering Theorem.
Similarly to that, it is not hard to prove, yet it changed our whole vision. Mine anyway!
It finally lifted the veil from our eyes about the real meaning of programs as data.
Programs as data was the key for Turing to show how "universal" machines exist that can simulate any others. This then implies machines that generate sets that have complements not generable by any machine. Unfortunately, always some diagonal tricks have to be used.
By Rice's Theorem, if we separate the programs according to how they run, then if one side is effective, the other can not be. It's that simple! Yet nobody realized it before.
The Formalist parrots regard this theorem only as a trivial consequence of results known earlier.
Or regard abstract versions of it. They don't want to deal with visions and understanding.
So then why did Rice discover this only in 1953? To answer this, is a taboo today.

Turing Falsifications and games

In the sixties in our math special highschool a very smart decision introduced computer classes. So the girls who couldn't get into uni went to be programmers. But that's about it! The name of Turing or what he discovered was never mentioned. And that was typical not just in Hungary.
Then much later, Turing became the "in thing". And a moron in America had a "vision" of rewriting history. He not only claimed that Turing was at once accepted by Gödel, which was important to claim because by this time Gödel was regarded as "God", but that "computability" is the perfect word for the new essence that Turing discovered. Beside sticking to Turing's obvious attachment to the computer word, it was very much tapping into the main stream because "home computers" became everyday household items. This name is paradoxical already and I wonder if a mother ever asked her teenage daughter: "Why are you always at that damn computer, when you have nothing to compute?"
Before the home computer age the "Turing Machine" name already became an "in thing" and was falsely assumed as the discovery of computers. Some have absolutely no idea about Effectivity, but like saying "Turing Machine", that indeed sounds "cool". I wonder if poor Turing's father had been called Smith, would these morons love to say "Smith Machine" too?
Computer, machine, derivation, generation, recursive were the words I used so far.
Now the mentioned moron wanted and succeeded to rename this recursive adjective to computable. The really illogical "enumerable" adjective in "recursively enumerable" didn't bother him so just translated it to "computably enumerable".
The way Turing used the word "computable" in the title of his famous article was unfortunate but made sense as I already emphasized this by them being a class wider than algebraic but narrower than strange. But really, the deeper problem is regarding computers as the only form of effectivity and so eliminate the word Effective altogether. This is by the way impossible by even those who stood in line because their articles are filled with this word. So why don't we just accept it as a perfect word for the bigger thing itself. Effective systems are much wider than computers. And games for example are the most ancient effective systems.
Some allowed steps are given and we apply them to win.
So Turing's discovery has to be true here too. And it is! In fact, it is possible to see the naturalness of non generable complements, without knowing about universal machines or understanding the breakthrough of Rice's Theorem, by simply regarding a game like chess.
Most chess players only want to win and so they never think about these following things:
We put some pieces on a board randomly and ask a good player "what is here?". Observe that for some situations the obvious reaction could be "this is impossible". For example, the pawns go only forward and start from the second line so they couldn't be on the first line.
But a more usual reaction is that "this is crazy". But then we can ask: "Could this be the result of a game, regardless how crazy the players were?". To establish this, we must find a game history, the actual steps from start. So to find this for a given setup can be quite hard.
But now I'll ask something more important that is not hard to answer at all.
Don't worry about individual situations! Instead, tell me which situations as a class is more complicated, the achievable ones or the not? And then the coin can drop! The achievable ones must be much simpler! Indeed, these each have a game history, a finite sequence of situations.
But to establish that a situation is not obtainable by a game, we must check all possible arbitrary long games and only then, after this infinite check can we tell that it didn't come about.
That's why the non theorems as a set is much more complex than the set of theorems.

Gödel, Robinson, Pressbuger

Some might say that I am an idiot because Gödel had the perfect vision just simply tried to be exact and avoided philosophical blubberings! Well, history proved this not to be true!
Namely, how with a better vision, he could have done much easier much more:
Omitting the induction axioms from Number Theory we get a much weaker "framework" theory like say algebra. But even as such, this system is almost useless because for example the seemingly trivial fact that a number can not be the consecutive of itself is not provable.
As a good side of this, this system is trivially non contradictory.
Indeed, in a contradictory system everything is provable.
This Robinson Arithmetic, in spite of its weakness is such that the non-theorems are non generable. This almost feels unbelievable but remember the logical axioms.
As I explained in my little detour at the Tait system, these mixing with even very simple logical rules can make the derivable theorems very complex. But the real magic of the Robinson system is that adding new axioms that are non contradictory, can not make the non theorem generable for sure. This is amazing because the non theorems will obviously shrink by adding new axioms. To see the truth rather than the road to prove it, is again better by the Effectivities.
At the undecidable statements of Gödel, the non generability of the non theorems is this truth.
There is a quite simple deeper layer of this truth if we go into the non theorems.
Some will be non theorems because they are the negations of theorems and so deriving these we would get instant contradictions and thus able to derive everything. These negatives of theorems should be called the "intentional non theorems". We don't want these to be theorems.
The deeper truth is then, that the theorems and the intentional non theorems are not selectable.
Meaning by this that there is no full selector of all statements that would include all theorems among say the left ones and all intentional non theorems among the right.
Such left-right full selector is nothing more than a generable set with generable complement.
Indeed, if both halves are generable we can just wait as selection which comes up.
Now we see how the Robinson Arithmetic can work even though its non theorems shrink by extensions. Simply because its theorems and intentional non theorems are not selectable.
Consistent extensions will keep these two inside of the wider theorems and intentional non theorems and thus also inside of the theorems and non theorems. So these two sets can not be a full selection. So if the theorems are generable then the non theorems can not be!
Just as at the simpler Gödel undecidablility, the actual proof of the truth goes back to the representation of effectivities by properties and finding tricky statements.
The fundamental new feature is "perfectly" representing the full selectors of the naturals.
This "perfect" means that not only we have two properties that represent the two halves by deriving the property cases for the set members but that the two properties are negatives.
Then indeed, non contradictory new axioms can not alter the old perfect representations!
So this perfect representation inherits to the new axiom system.
And then a "simple" argument shows that a representation of the full selectable sets doesn't allow both the theorems and non theorems to be generable. Here I wasn't sloppy saying "that a representation" and left out the perfect condition because this condition is here irrelevant!
So the perfect representation was only necessary to be able to inherit.
Now I show this "simple" "unselectability" argument but to see its finesse, first I show the old "undecidability" argument too that Gödel used.
We'll also redefine the basic concepts involved, so let T be a set of statements.
We have the set of theorems in mind but we don't assume this here yet.

T is trivially inconsistent if there is an A that both A and $\neg A$ are in T.
T is complete if for any A exactly one of A or $\neg A$ is in T.
T is incomplete if it is not complete and this can mean two things:
Either it is trivially inconsistent or not but there is some A that neither A nor $\neg A$ are in T.
We call such A , $\neg A$ pair as an undecidable pair in T.

Undecidability Argument:

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

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

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

Indeed:
Using $\neg Q$ as P

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

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

A crucial feature of the T set of theorems above was ignored. Namely, that they are generated!
Now we regard all the generable number and statement sets as a given G set.
If for an S number or statement set both S and its complement $\neg S$ are in G then we call S and $\neg S$ as fully selectable but now in these followings just in short as selectable.

Unselectability Argument:

If G is such that:

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

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

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

then T can not be selectable.

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

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

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

The first line gives what we had in our previous argument and so it only proves that our assumption of a selectable T implies incompleteness but this doesn't refute the assumption.

On the other hand, the second line with $P = R$ gives:

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

which is impossible.

All this throws light even better on the role of generability.
We may add as “new axioms” to the Robinson system all truths among the naturals.
Then the unselectability argument just says that the truths or falsities must be non generable.

In fact, they will be both non generable as can be seen by other arguments.
But for obvious new axioms that are generable, like the induction axioms, the theorems are trivially generable, so we get that the non theorems are non generable.
So undecidability or the infamous "Incompleteness" of the full Number Theory comes out from a sub system where undecidability is trivial. The non selectability of the theorems and intentional non theorems of this sub system inherits to the full system. And there it forces a non trivial undecidability. This shows clearly that not undecidability is the real point.

This makes also perfect sense of a historical prelude to Gödel. Just a year before Gödel realized the undecidable statements in arithmetic, Pressburger proved that without multiplication, this Baby Arithmetic is complete. It can decide every statement.
But his proof is not trivial at all. So even this Baby Arithmetic is not trivially complete.
Above I said that the infinity of logical axioms explains how the derived theorems can be complex in a seemingly simple system too. So complex that the complement is not generable at all by any means. But "explains" doesn't mean that it always causes it!
Pressburger result shows that the potential complexity can be avoided and have generable complement. Thus the opposite case, that is actual complexity so that the complement is not generable, is not visible from the forms of a derivation system. Only the representabilities shows this. So in a sense, we must admit that all was there in Gödel's original article.
Because it contained the crucial points that made multiplication the extra step causing the representabilities.

Peano Rules and beyond

To see this most clearly, we should avoid the successor function $S(x) = x + 1$ used as basic symbol and rather stick to relations. So then $x \lhd y$ denotes consecutiveness.
The price is that we can not get the values of $1 , 2 , \ . \ . \ .$ as $S(0) , S(S(0)) , \ . \ . \ .$ So we have $1 , 2 , \ . \ . \ .$ as infinite many basic names with the infinite many name axioms: $1 \lhd 2 \lhd \ . \ .$
The additional axioms to describe $\lhd$ and to define the operations bear Peano's name.
An inclusion of 0 as natural number was also accepted by Peano to simplify his rules and there is an anecdote about this: 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".
I agree with his wife and will not include 0 as natural.
The three claims about $\lhd$ are very "simple":
For every x there is a single y that $x \lhd y$.
For every y except 1 there is a single x that $x \lhd y$.
There is no x that $x \lhd 1$.
Why I put the simple in quotation mark, will be explained after I explain Gödel's crucial idea.
But back to Peano, he realized that the arithmetical operations can be "defined" as:

$$x \lhd z \ \rightarrow \ x + 1 = z \quad , \quad x + (y +1) = (x + y) +1$$

$$x \cdot 1 = x \qquad , \quad x \cdot (y + 1) = x \cdot y + x$$

$$x^{1} = x \qquad , \quad x^{(y+1)} = x^{y} \cdot x$$

Or regarding the operations as relations and avoiding the functional tricks in the second axioms:

$$x \lhd z \ \rightarrow \ x + 1 = z \quad , \quad x + y = z \ \wedge \ y \lhd y' \ \wedge \ z \lhd z' \quad \rightarrow \quad x + y' = z'$$

$$x \cdot 1 = x \qquad , \quad x \cdot y = z \ \wedge \ y \lhd y' \ \wedge \ z + y = z' \quad \rightarrow \quad x \cdot y' = z'$$

$$x^{1} = x \qquad , \quad x^{y} = z \ \wedge \ y \lhd y' \ \wedge \ z \cdot y = z' \quad \rightarrow \quad x^{y'} = z'$$

So the operations are now three variable relations each "defined" by two axioms.
The first is an initial statement, the second an implication that really makes us able to get the cases for all names by using the name axioms too.
To derive $4 + 3 = 7$ for example we go like this:
$4 + 0 = 4$ by the initial axiom, $0 \lhd 1$ and $4 \lhd 5$ are name axioms so using the implicative axiom,
$4 + 0 = 4 \wedge 0 \lhd 1 \wedge 4 \lhd 5 \rightarrow 4 + 1 = 5$ we get $4 + 1 = 5$ by the Logical rule Modus Ponens.
Then $1 \lhd 2$ and $5 \lhd 6$ will give $4 + 1 = 5 \wedge 1 \lhd 2 \wedge 5 \lhd 6 \rightarrow 4 + 2 = 6$ and so $4 + 2 = 6$.
Finally $2 \lhd 3$ and $6 \lhd 7$ will give $4 + 2 = 6 \wedge 2 \lhd 3 \wedge 6 \lhd 7 \rightarrow 4 + 3 = 7$ and so $4 + 3 = 7$.
For multiplication cases of course we need to derive the needed addition cases and for exponentiation both addition and multiplication cases.
We get all true operational cases and so the "definition" was perfect but we still had to use quotation mark because definition as such means explicit definition, using earlier defined or accepted basic relations only. Here the crucial implicative statements had the relation in the condition already. So we actually had a circularity. That's why the usual name is Peano rules.
Which of course doesn't really explain that these are perfectly okay axioms and only to call them defining axioms is confusing. Indeed, a definition should be explicit so not containing the defined relation in the conditions too. Axioms of course can be implicit.
Quite amazingly, if we want to continue our operations infinitely by similar rules, the power of general recursion allows to get all those as a number dependent $x\ [n]\ y = z$ relation.
So $n=1$ should give addition, $n=2$ multiplication and so on. The rules are then:

$$x \lhd z \quad \rightarrow \quad x\,[1]\,1 = z$$

$$n \lhd n' \quad \rightarrow \quad x\,[n']\,1 = x$$

$$n \lhd n' \;\wedge\; x\,[n']\,y = z \;\wedge\; z\,[n]\,x = z' \;\wedge\; y \lhd y' \quad \rightarrow \quad x\,[n']\,y' = z'$$

The first rule simply says that for $y = 1$ value, addition is same as $x \lhd z$.
The second says that for the other $[n]$ operations the initial value is always x.
The third rule says that to step in y, we need not just the value of the present n' operation but the previous n too, applied for the old result z with x.
And indeed, to increase the multiplication value from y to the next y' we just have to add x.
To increase the exponentiation value from y to y' we must multiply by x and to increase the $[4]$ value from y to y' we must raise the old result to the x power.
So $x\,[4]\,1$ is x by the second rule and then $x\,[4]\,2$ is x^x.
Then $x\,[4]\,3$ is $(x^x)^x = x^{(x^2)}$ and we shouldn't write simply x^{x^x} for this because the $x^{(x^x)}$ value is different and could be understood for it just as well.
This is a consequence of exponentiation not being arbitrary in its order unlike addition and multiplication. So, for multiplication we can simply say that it is repeated addition and for exponentiation that it is repeated multiplication but for $x\,[4]\,y$ we shouldn't just say that it is repeated exponentiation.
A simpler fact is that the exchangeability of order also fails for exponentiation.
Indeed, $2^3 = 8$ but $3^2 = 9$ which feels trivial but same rules created addition and multiplication and for those the order is irrelevant. This is a mystery!
Observe something else that's interesting! For the old normal operations the operational value z was always bigger than the x , y input values. Now with $x\,[n]\,y$ this fails.
For example, for all n values we get merely x as initial value at $y = 1$.
This fact that we can get small z values for big n hides the fact how fast z grows.
This caused big havoc at the early recursion period of Effectivity.
Then the normal operations were first generalized in a different direction as f functions that are given initial values and then defined for step by step increasing inputs.
These primitive recursive functions can not grow as fast as our $x\,[n]\,y$ and this was regarded as a complexity of $x\,[n]\,y = z$. The truth is quite the opposite.

Our rules only give the impression that the collectable tuples are complex because we have the freedom to choose derived cases and stupid choices can become an infinity of cases that are only a very small subset of all derivable cases. But we can be much smarter too.
We can regard an m fixed value and try all numbers up to m as inputs that is case conditions in all our finite many defining conditions. Obviously as start we'll only be able to use these numbers in the ⊲ basic relations there. Then we get some target tuples and we again only regard the ones using values up to m. These can now be used in our second try of all conditions. We again get new tuples with values up to m and we use these again. Repeating this, we get a stage where no more new targets can be obtained. Simply because we couldn't get infinite many target tuples by using only numbers up to m. So we derived all target tuples with using values up to m.
Observe that this m will also be the maximal value in all of our derivable targets above.
Indeed, in every target in our rules above, the maximal variable value was always at least as big as other values in the conditions. Namely, the maximal was always z' except at x [n'] y' = z'. And here if n' is the same or bigger than z' then the only variable not up to z in the conditions is n.
m appearing as maximal in the targets then implies that a target with values up to m can not come about by increasing m to an M. Simply because then M will be the maximal.
This means the same that the tuples not in our targets up to m will remain outside tuples as we increase m. So we can derive the outside tuples too as follows:
We derive the inside ones in the above systematic manner using m = 1 , 2 , 3 , . . and at every such m value once we have all the derivable ones, we check all tuples formable up to m and list the ones not derived.
We can get rules that avoid this maximality of the targets by simply dropping out variables.
But such rules will still not produce a relation that trivially has no generable complement.
To see why, we must remember Rice's Theorem. We must collect programs!
Generating statements by Logic can be more complex because due to the representability of effective methods within, we can actually collect programs.

Gödel really

This would bring us back to logical derivations but I still didn't show Gödel's crucial discovery, neither the mentioned deeper problem behind the simple first three axioms of ⊲.
Amazingly, for Gödel's discovery we must go opposite to the generalization we "wasted" so much time just now. In fact, we must omit even exponentiation. So we need only seven axioms beside the infinite many name axioms. Three for ⊲ and four for addition and multiplication.
This is Robinson's system that as I said could have been also enough for Gödel's discovery.
Suppose we have a $(t_1 , t_2 , . . . , t_n)$ so called tuple of natural numbers!
Can we create a code for it? We instantly think of a single number but a pair of M , N would be just as good. Indeed, for arbitrary long tuples just a double is a perfect result.
The more important requirement is that we should be able to recover our tuple from M , N using multiplication with logical symbols. This luckily allows division or rather dividability and even remainders. And indeed, with these it is possible to define an explicit T(M , N , i) = t expression which for any fix M , N values will be true at a single t for some i values that always go up to n for sure. This is great because this way we don't need to decode n, rather just go up to n many members in our sequence of the defined t-s.
Now let's see what such expression can do because it is unbelievable.

We can explicitly define exponentiation! Indeed:
$x^y = z$ means that there is a $(t_1 , t_2 , . . . , t_y)$ tuple that $t_1 = x$, $t_2 = x^2$, . . . , $t_y = x^y = z$.
Here we still have the dots but we can avoid them by saying that:

$$\exists M \exists N \{ T(M,N,1) = x \wedge \forall k [k < y \rightarrow T(M,N,k) \cdot x = T(M,N,k+1)] \wedge T(M,N,y) = z] \}.$$

We can see that any other repeatedly calculated tuples can be formalized if the calculation is expressible by addition and multiplication. And indeed, by such tuples we can describe anything effective. So not just exponentiation but all generable collections just became explicit.
Then we can show that the cases of these explicit definitions are even derivable in our system.
So the generable sets are not just explicitly definable in our language but representable in our axiom system.
The theorem we need for the concrete $T(M,N,i) = t$ is called the Chinese Remainder Theorem:

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

The notation reveals that our tuple will be the remainders in N but how do we get the mentioned s values from a single M.
Let $t_1 + t_2 + \dots + t_n + n = m$ and $2 \cdot 3 \cdot \, . \, . \, \cdot m = m\,! = M$
Then $s_i = i\,M + 1$ will satisfy their conditions in the Chinese Remainder Theorem.
The first condition, the t values being under them is trivially true because :
$t_i < m < m\,! = M < i\,M + 1 = s_i$
The second condition, the relative primness means that:
For every $i \leq n$ the $i\,M + 1$ values are all relative primes to each other.
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 $(k - j)\,M$ too. p divides separately and can't divide M since it is an M divider.
But neither can divide $k - j < n < m$ because $M = m\,!$, so $k - j$ is an M divider too.
Thus we can use the theorem and claim that such N exists.
This is a concrete number, though the theorem doesn't give it explicitly. For our purpose it is enough that such N exists. The explicit expression for $T (M , N , i) = t$ is :

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

This 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.
Then the expression is giving a full infinite sequence of t-s.

For a proof sketch of the Chinese Remainder Theorem observe these:
Let $P = s_1 \, s_2 \dots s_n$. The $0 , 1 , 2 , \dots , P - 1$ 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.
There is an uglier method to find a $T (M , N , i) = t$ expression without the Chinese Remainder Theorem and it is using an N that is explicit from the tuple.

Non Standard Models

Now I can come to the mentioned deeper problem behind the seemingly simplest part of our axiom system, the three axioms of ⊲. The problem starts by first realizing that they don't claim what we think they do, that is describing: $0 \lhd 1 \lhd 2 \lhd \ . \ . \ .$

To see this, imagine that we add to our $0 , 1 , 2 , \ . \ . \ .$ names infinite many new ones denoted as: $+1 \ , \ -1 \ , \ +2 \ , \ -2 \ , \ . \ .$ This doesn't reveal much but put them under our real naturals as:

$$0 \lhd 1 \lhd 2 \lhd \ . \ .$$

$$. \ . \ . \ \lhd -2 \lhd -1 \lhd +1 \lhd +2 \lhd \ . \ . \ .$$

Quite unbelievably, our three axioms remain true for the totality of these objects
Every number has a unique next.
Every number except 0 has a unique previous.
Only 0 has no previous.
What a disaster! But the worst comes now! We might think that accepting the other axioms somehow these crazy numbers might disappear. Not so! They only become more complicated.
Now the completeness of our logic claiming the derivable statements to be true in all realities makes more sense and also the undecidability result of Gödel. In fact, we might wonder why not such non standard realities were found having some non standard truths in them, which then were obviously not derivable. But most surprising is that these realities were known twenty years before Gödel found his method to show non derivable statements.
That's how mysterious these non standard realities are. They don't reveal their alternate truths.

Church Thesis

Now we can return to Turing's famous article that beside the introduction of machines aimed already in its title to disprove Hilbert's mentioned dream for deciding the logical necessities.
He didn't claim it as a proof quite explicitly, and this is the point! Explicitly he couldn't!
And that's also the reason why the computability renaming of everything that is effective, is crazy! We do not know what effectivity really means! We have examples or rather frameworks of effective systems. Recursive functions, Turing machines, Machines with memory registers, Games, and of course the derivation methods of different logics are all such frameworks.
We can see that these are able to imitate each other so we claim as Church or Turing "thesis" that behind all lies a universal Effectivity.
I don't want to blubber about this any more because a much more important fact has to be told.

Randomness again

Namely, that the other such "outside" concept "in" mathematics that emerged is "Randomness".
I already went ahead already twice in this direction but didn't start with the simplest!
We tend to believe that if we flip a coin infinite many times then the fact that it has half half chance for each side, somehow forces that sooner or later both must happen.
Murphy's Law says that the bread always falls on the buttered side.
This belief quite oppositely, just tries to tell the "inescapable" reality, that if you keep on dropping your bread then once this messy possibility will happen.
In truth, it is only inescapable in a life of badly repeated trials.
This sounds mystical so something even more mystical:
We all share the common human intuitions as a priory eternal foundation of mathematics.
Most mathematicians and philosophers at present would argue with me.
Luckily, about this I will never argue. What I want to express though is that I am not as idiotically simple minded as my belief would suggest!
In my view these a priory eternal intuitions are hidden behind the material experience.

We must dig them out from our souls through rejections of falsities that seemed plausible through our ignorance. One of the deepest such rejection in my life was to finally realize that the above mentioned certainty of a coin landing on head if we try it long enough, is false!
A coin can land on tail forever and only our limited experience of such infinite trials suggests otherwise! Since this admission was a dramatic moment, I will be personal for a bit longer.
I was still a believer in Randomness four years ago when I came back to Hungary for a visit after twenty years living in Australia. An event right at the start was the wedding of my brother's son. I was also to meet his grandson from his daughter. This boy was claimed to be a mathematical genius and of course I was skeptical and said "we'll see".
At the dinner we talked all the way through and I was very impressed. The other daughter of my brother was listening who was a non practicing math teacher. As it turned out she was also very impressed by how I was able to "widen" the boy's mind. I don't know how coin flippings came in but she said that she doesn't understand why some outcomes are labeled as random when all outcomes are just sequences of head and tail. I forcefully explained that the half chance, like any chance, even the smallest, requires that sooner or later it has to happen.
Something happened in that "forceful" but empty argument that changed my intuitions.
Only months later did it become clear to me that there is absolutely nothing that could force a coin to fall on head. The mathematical possibility of all tails is a correct intuition and our expectation of "finally" a head happening is a false one.
This of course meant to rethink all that I knew about Randomness.
In high school I discovered the earlier mentioned solovay randomness without even attempting any other approaches. It was almost a trivial faith that the infinite or non infinite occurrence of some beginning properties should decide if a sequence is strange or random.
The finite total chance of the beginning property is identical with the property diminishing fast.
This was my main dilemma that this fast diminishing is such a subtle difference from normal diminishing.

$\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \; . \; . \; . \; . \; .$ is only slowly diminishing because the sum is infinite.

$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \; . \; . \; . \; . \; .$ is fast diminishing because the sum is finite, namely 1.

In the sequences themselves the difference is hardly visible and yet . . !
Now comes the big surprise! When I asked people about concrete beginning properties to be stopping or continuing forever then they always felt the right way.
Feeling the stop necessary only for fast diminishing properties.
Strangely, this whole line of fast diminishing properties stopping in random sequences came from the strange property of a beginning being repeated or mirrored or mirrored oppositely.
The Goldbach Conjecture claims the impossibility of this last ever happening if instead of the random 0-s and 1-s we use these to mark the composites and primes.
Indeed, $2n$ to be the sum of two primes means that n is the average of the two and so not having such primes means that mirroring the primes under n we never hit one above, that is the $0 , 1$ values are antisymmetrical to n as center.
So I realized that the Goldbach claim is merely a universalization of the expectable by chances.
After a value the beginnings have to be such by sheer chances and the "value" here being from the start is merely a coincidence.
I still believe in this, that is not seeing any deeper importance in the adding of the two primes.
Instead, believing that primes behave in random manner to some simple tests in general.
Then Von Mises or Ville were unknown names to me too, just as Turing.
And to my fault, I only returned to Randomness forty years later.