Cantor

We explained his grand recognitions of the sets in "The f-widening Theorem" article.
We also showed the generalization of his anti diagonal method to see that the P set of all subsets of an A set are bigger than A. Amazingly, the Axiom Of Choice was not involved in this while a generalization of his simpler result, that the fractions are sequencable will need this. Both results can be formulated in abstract form by the use of set exponents, in particular with A^2 and 2^A. This first means the set of ordered pairs made from A and so directly correspond to the fractions if A is the set of naturals. The second will correspond to the set of subsets of A that we earlier called as the P power set of A. By the way, we used this P power set to make a c choice function from it, by which we defined the complementing f continuation function, by which the f-widening Theorem gave the Well Ordering Theorem.

To make a common meaning for these two new abbreviations is to regard a B^A "set power" as the set of choice functions defined on A and making choices from B.

So $\{0 , 1\}^A$ is then the set of all possible binary choice functions from A.
The 0 and 1 values could be head and tail and thus we would get the possible outcomes for coin flips at all a elements of A. An other similar duality that jumps in mind is of course yes and no choices and if regard these as collection decisions for every a element of A, then indeed we get all possible subsets of A. So practically $\{0 , 1\}^A$ becomes 2^A.

Observe also that $\{0 , 1\}^\omega$ corresponds to the binary "decimals".

$B^{\{0,1\}}$ is the possible choices from {0 , 1} with b elements from B as values.
This actually means ordered pair formations from B.

So for example $\omega^{\{0,1\}}$ practically means the fractions.

These observed meanings match the general claims that $B^2 \sim B$ while $2^A > A$.

The simpler but harder to prove $B^2 \sim B$ fact will have two amazing consequences that continue the easier provable $2^A > A$.

The first such consequence is quite shocking because 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^A is by simply using as B a set that is already bigger than 2^A.

Clearly, if $B > 2^A$ then $B^A > 2^A$ too and we claim that this trivial way is the only way to go above 2^A. So $(2^A)^A \sim 2^A$. Simply because: $(2^A)^A \sim 2^{(A^2)} \sim 2^A$.

An even more surprising consequence is that actually:
$(2^A)^A \sim A^A \sim 2^A$ because $(2^A)^A \geq A^A \geq 2^A$.

The other amazing consequence of $B^2 \sim B$ will be more involved and relies on the concept of the "minimal" well ordering of an A set. Since in a well ordered set every property has a first case, thus we must have a first B beginning of A that $B \sim A$. If such is only A itself, so A has no equivalent beginning then we say that A is minimally well ordered or it is a cardinal.
Then any other well ordering of A can be regarded as a stretching of A.
The possible longer and longer stretchings of an A are all the well orderings before the next size where a stretching of A 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 the naturals.
Their combined total well-ordering is abbreviated as ω_1 because ω is used to abbreviate the naturals and also the cardinal as the minimal well ordering of the sequencable sets.
ω_1 is again a cardinal which again can be stretched as far as possible giving ω_2 and so on.
The story doesn't stop because combining the ω , ω_1 , ω_2 , . . . well orderings we get a well-ordering ω_ω that actually has to be a new size because it can't be equivalent to the earliers.
And it is again a minimal well-ordering of that size so it is a cardinal.
So this ω_ω came about not as next infinity by stretchings rather as a limit of all earliers.
And the minimal well orderings or cardinals continue after ω_ω as $\omega_{\omega+1}$, $\omega_{\omega+2}$, . .
And then again we will get limit formation cardinals and stretching formed ones again.
So the cardinal indexes can be all kind of well-orderings or also called as ordinals.
But this first impression of boring monotony is false! Deep puzzles lie behind the monotony of the ordinals already and an extra layer behind the cardinals!
For example, the above duality of the stretching cardinals and the limit cardinals can be attacked by the observation that while ω_ω has a simple ω subset that goes all the way, the stretching formed next cardinals don't seem to have such.
But strangely the first infinite cardinal ω is a limit cardinal but also has no smaller subset that goes all the way. So is it possible that an ω_α is again a limit cardinal and yet has no smaller subset going all the way? And presto, we arrived at an undecidable problem.

Less important problem is the notation of ordinal formations, or ordinal arithmetic.
$\alpha+\beta$ is simply a continuation and $\alpha\,\beta$ means α type with the elements replaced by β types.
Repeated multiplication can be abbreviated with exponentiation like $\alpha\,\alpha = \alpha^2$ and this doesn't cause yet any confusion with set exponentiation.
But combining ω , ω^2, ω^3, . . . types we get a new type that should be abbreviated as $(\omega)^\omega$.
In fact, to be more visual we can even define this as $(\omega)^\omega = \omega + \omega^2 + \omega^3 + \;.\;.$
This is just a sequencable set, nothing to do with the ω^ω set power.
$(2^\omega)^\omega \sim \omega^\omega \sim 2^\omega$ by the earlier mentioned amazing irrelevance of the base but these are now not well orderings so cardinals, rather sets made from ω as the set of naturals.
But beyond this bit of confusing ambiguity the biggest real puzzle of Set Theory is that we have no idea where this infinity, also called as the continuum lies among the ω_α cardinals.
The only "good thing" is that we know it can not be ω_ω. We come to proving this soon but to have a better vision of the well orderings, we should find some other conditions about an α well ordering being jammed towards its end, beside being minimal, that is being a cardinal.
The simplest such condition is that if we can cut off any proper beginning from α the left over end still remains similar. ω is obviously such cuttable but $\omega + 5$ or $\omega+\omega = 2\,\omega$ is not.
Then $\omega+\omega+\omega+\ldots = \omega\,\omega$ is again cuttable.
And actually, after every cuttable α the next is $\omega\,\alpha$.
So being cuttable is a boring monotony again. Surprisingly, these boring cuttable well-orderings are the way to prove that $\alpha\,\alpha = \alpha^2 \sim \alpha$ which then implies $B^2 \sim B$ for sets.

The crucial new idea is what we hinted already above, to regard not merely beginnings that we can cut off rather 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.

As we could guess from our earlier mentioned "hints", this concept relates to the cardinals. Strangely, they almost coincide and we even hinted about the imperfection.

We start with the promised second use of $B^2 \sim B$.
$\alpha^2 \sim \alpha$ implies that if after an α cardinal the next one 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 cardinals but the being next condition doesn't follow so it leaves open the question whether a limit cardinal could be a cofinal.
As we mentioned earlier ω is such and it means that it is the combining of so many cardinals as it is itself. Which is trivial because the naturals are the finite cardinals.
But the existence of an ω_α that is again such but with infinite cardinals combined, can not be derived from the axioms of present Set Theory.
By the way, you can find the just mentioned "quite easy" proof in Appendix 5. of "The f-widening Theorem" article.

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 $2^A > A$ generalization of Cantor's non sequencability of the decimals. In fact, the whole idea of the B^A set as functions can be "generalized" first by allowing different B-s as the a elements of A itself.
Then the functions are actually the choice functions from A and their set is denoted as $\prod A$.
So we can claim that if all a elements have at least two elements, that is $a \geq 2$ then $A < \prod A$.

I used quotation mark above and indeed, this is a bad generalization. It doesn't say more rather less because this would follow from $B^A > A$ anyway and we only restricted the A set.

This bad idea was perfectionized by König allowing to make choices not directly from the a elements of A rather some assigned $k(a)$ sets. Then with $k(a) \geq 2$ we could increase the size of not A rather R_k the range of k. But this wouldn't be anything new again.
König realized that instead of the primitive $k(a) \geq 2$ condition, an opposite $k(a) < a$ can be made and then increase the size of not R_k rather its $\bigcup R_k$ combined set.
The jump is still the $\prod A$ set but now A 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 a elements of an A 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 a elements of A as T-s.
Namely, we should regard the k that assigns $\{a\}$ for every $T = \{(a,0),(a,1)\}$ pair made for each $a \in A$.
$k(T) < T$ because $1 < 2$ and $\bigcup R_k = A$ and also $\prod D_k \sim \{0,1\}^A$.
Indeed, every element of $\{0,1\}^A$ 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 $T_1 \subseteq T_2 \subseteq T_3 \subseteq \ . \ .$ sets with $T_1 < T_2 < T_3 < \ . \ .$ let's regard: $k(T_{i+1}) = T_i$.

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

We can apply this for the $\omega < \omega_1 < \omega_2 < \ . \ .$ minimal well orderings where $\bigcup \omega_i = \omega_{\omega}$.

So we get that: $\omega_{\omega} < (\omega_{\omega})^{\omega}$. But König also proved that $(2^{\omega})^{\omega} \sim 2^{\omega}$.

So then indeed, $\omega_{\omega} \sim 2^{\omega}$ is impossible.

This fact that König also proved, we already mentioned in general form as $(2^A)^A \sim 2^A$.

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.

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 actually the whole infinite dimensional space has also only same many points as any tiny 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 Zeno and maybe even to the mystery of where the cardinality of 2^{ω} lies.

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 first prove $B \geq 2 \rightarrow B^A > A$:

$B^A \geq A$ is trivial so enough to show that $B^A \sim A$ is impossible, that is if g is a function defined in A and having values in B^A then some $f_{-g} \in B^A$ is missing as g value.

Define $f_{-g}(a)$ as anything but $g(a)(a)$. Since B has at least two elements, this can be done.

Then $f_{-g}(a) \neq g(a)(a)$ for every a, so $f_{-g} \neq g(a)$ for every a 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_{-g} 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 T_g subset of T can maximum have as many elements as $k(T)$.

So $T_g \leq k(T) < T$ and so $T - T_g = T_{-g} \neq 0$.

Make f_{-g} by picking for every T an element from T_{-g}.

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

But $f_{-g}(T) \notin T_g$ and so $f_{-g} \neq g(r)$.

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 the k(a) sets assigned to the a elements of our A set.
So $f(a) \in k(a)$ and these are the ones that we call now choice functions on A.

If all k(a) are the same B set then we get back B^A 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(a) 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 in Logic.
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.
It is best not to regard A as a fix 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(a) values as possible outcomes of experiments carried out at every a 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 A as windows.
If we make choices for each a from k(a) but only for a w window then this is called a v valuation of w. From v itself its window is abbreviated as [v].
The word "valuation" suggests that we make the choices.
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 S subset of A and there it is denoted as f_S. 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 A 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 A 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:

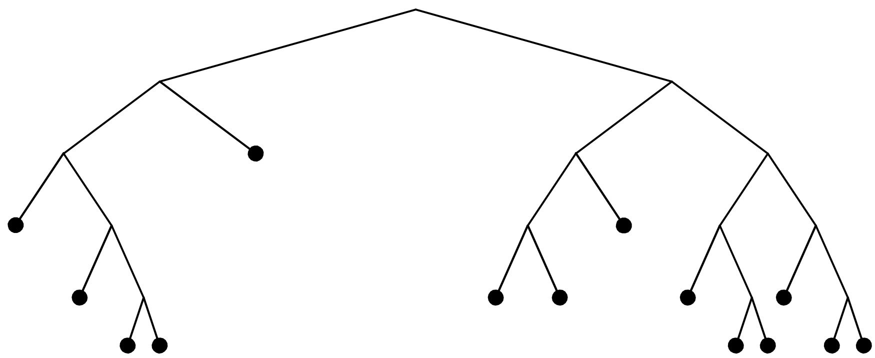
If a V prediction set is sure, then already some finite subset of V is sure.
If every finite subset of a V taboo set is avoidable then V is avoidable too.

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 A 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 A anyway because our construction goes through all elements of A. So we imagine A 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 A 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 a_α?
That choosing any window after a_α we can evaluate it so that this added to the choices up to a_α including the new choice for a_α, we find no taboo.
This indeed implies taboo avoidance up to a_α by simply regarding empty window after a_α.
But can such choice be made for all a_α? Suppose a_α is the first where this couldn't be done.
This means that for every choice of a_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 a_α this set is again finite so is a window.
Finally, lets add a_α to it too. Every evaluation of this, added to the beginning before a_α will contain taboo. Indeed, if one would not contain taboo then checking what value a_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 a_α.

This at once shows that if a_α is a_1 or a later member but with a previous member, then such failing of choice for a_α couldn't happen. For a_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 a_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 a_α. But then this earlier member couldn't have been chosen.

König Path existence

It's very educational to look at the simplest case of the taboo avoidance.
The A set is the naturals and we have binary choices. So, we regard the binary sequences.
The w windows are finite sets of naturals. The v valuations are assignments of such w with the possible 0 or 1 values. And a V taboo set is a set of such v assignments.
Our compactness result says that if on all finite subsets of the naturals we can avoid V then we can avoid it on all naturals too. So there is a taboo avoiding infinite binary sequence.
The avoidability on all finite sets is then equivalent with avoidability on all beginnings.
And more importantly, having such B set of taboo avoiding beginnings, dictates a strategy to build a taboo avoiding s sequence without worrying about the taboos at all!
So, w , v , V are all thrown out the "window".
All we need is that our s sequence should "sub-continue" in B.
Meaning, that all beginnings of s should be a sub beginning in a b in B.
Indeed, then no taboo can appear in s because it would have to be in b but B avoids such.
So if we can show that for any infinite B set of beginnings there is an s sequence that sub-continues in B then the taboo avoidance follows as triviality.
The proof of existence for such s is amazingly simple. We build s digit by digit.
The first choice, 0 or 1 goes by simply picking one that is the start of infinite many beginnings in B. One of them must be since otherwise B would be finite.
If you want to be more definite then choose 0 if it is the start of infinite many beginning in B and chose 1 if 0 was not. Say we chose 0. The continuation is again two kind 00 or 01.
Now again one must be in infinite many among those that started with 0.
And so on, we can always chose an infinitely continuable beginning and thus these choices can never get either to a dead end.
The name of our result comes from a visual representation of the binary beginnings as an infinite "tree", though better envisioned upside down:



Going left or right means 0 or 1. The black dots are the b beginnings.
In our picture above the shortest so upmost black dot represents 01 and the two next lengths ones are 000 and 101.
In our picture the black dots all terminate which means that it represents a non continuing set of beginnings.
If a beginning in B is sub beginning of an other one in B , or saying it differently, the other is continuing the first, then the trip to the second black dot will go through the first.
In our picture this never happens and so our picture shows a non continuing B.
The condition that B is infinite, simply means that we have infinite many black dots which implies that we must have ones arbitrary deep down.
The claim is that we have an infinite path downward, going through the marked branchings.
Going through black dots is not necessary and in a non continuing set as the picture shows above, it is definitely impossible. But just simply avoiding the black dots is obviously not enough to avoid a later non continuing black dot or "dead end".

The simple but ingenious trick is to always go where the further continuations have still infinite many black dots. Indeed, if we use this strategy from the start then where we are at a junction we always have still infinite many black dots downward in the two directions together.
So in at least one of the continuations there have to be again infinite many black dots further.
Simply because two times finite would be only finite together.
If both of them have still infinite many black dots ahead then of course we can choose either.
This way, an encounter of a single black dot is excluded too.
The discoverer of this abstract fact was the son of the earlier mentioned König.

An interesting representation of the “spirit” of the trick used, is to regard the junctions as “Y” sections of roads in a forest guarded by fairies. To make our decisions to go left or right we can ask the fairies. But obviously we can not ask “Which way to go so I can go forever?”.
Indeed, that would be asking something about the result of our own decisions.
So we must ask something objective that exists in the forest already.
And of course the question should be: “In which direction are there infinite many fairies?”.
There is an other less mythical visualization of the beginnings that nevertheless makes the path existence very plausible.
Here the sequences are regarded as points of [0,1] located by the infinite binaries, corresponding to the sequences, exactly as the infinite decimals locate them by ten divisions.
This would suggest that the finite binaries are regarded simply as omitting the infinite many zeroes from sequences ending this way and so representing the halving points.
But there is a much better, heuristically new representation!
Namely, as the closed halving intervals!
The one and two long beginning’s representations are as follows:

$$0 \approx [\,0\,,\tfrac{1}{2}\,] \qquad , \qquad 1 \approx [\tfrac{1}{2}\,,\,1]$$

$$00 \approx [\,0\,,\tfrac{1}{4}\,] \quad , \quad 01 \approx [\tfrac{1}{4},\tfrac{1}{2}] \quad , \quad 10 \approx [\tfrac{1}{2},\tfrac{3}{4}] \quad , \quad 11 \approx [\tfrac{3}{4}\,,\,1]$$

This resolves the problem that the full 0 ending sequences have an alternative full 1 ending version too.
This corresponds to the problem among the infinite decimals, we learned already in Elementary School, that infinite many zeroes at the end have an other version with infinite many nines.
But now for our purpose something much deeper is useful, namely that:
For an s sequence to sub-continue in B is equivalent with the P point representing s to be approached by the intervals representing the beginnings in B.
And indeed, infinite many intervals must approach a point.
Observe that these intervals may be such that none is inside an other, corresponding to a non continuing B set of beginnings. The existing limit point or path is still true. Then this path will not be a continuation of any beginning in B or in short, it will not “continue from” B.
Correspondingly, the limit point will not be “covered” by any of the intervals.
Strangely, in spite of this visual plausibility, an exact argument in [0,1] would have to go back to sequences. Or use Cantor’s Common Point or Dedekind’s Continuity principle.