Comparing Infinites

1. An S set is subset of T or narrower than T or T is a superset of S or wider than S or in abbreviated form $S \subseteq T$ if all elements of S are elements of T too:
$e \in S \rightarrow e \in T$. Or more precisely, using the $\forall$ quantor for the "all" or "every" condition: $\forall e\,(e \in S \rightarrow e \in T)$. This of course allows that $S = T$ that is they have the same elements. That's why the $\subseteq$ symbol "contained" the equality.
By claiming that S is a proper subset of T, we mean that there are elements of T that are definitely not in S and we denote this as $S \subset T$. We could say that then S is "smaller" than T but we won't because we want to use this word for a true quantitative comparison. So, instead:
An S set is smaller than T or in abbreviated form $S < T$, if for any f function defined on S and having values in T, the f values can not take up all the T elements.
So, for any f we define on S with having values in T, there is at least one t element of T that is not taken up as value for any s element of S, that is $f(s) \neq t$.
Of course, the part of our assumption that f only takes up values in T can be omitted because the impossibility for those that stay inside T would make it also impossible for others and in reverse obviously too. To avoid the condition of being defined only in S can not be omitted because other f values defined on somewhere else could always take up the full T. For example by using the T elements assigned to themselves. Yet actually even this condition of being defined on S can be avoided. First of all, if we only know that f is defined in S but not necessarily on S, that is not for all values, then this makes the impossibility stay. Now if f is defined anywhere then we can still simply regard the values only in S and claim about these that they can not produce all T elements. So:
For all f functions, there is a t in T that for all s in S : $f(s) \neq t$.
Abbreviating the "there is" condition as $\exists$ and negation as $\neg$ our definition becomes:
$S < T \,:\, \forall f\ \exists t\ \forall s\ [\,(t \in T) \text{ and } (s \in S \rightarrow \neg\, f(s) = t\,)\,]$.
We can use $f(\ )$ not just for the assigned values to values placed into ( ) , but for sets of values and giving the sets of the assigned values. So $f(R)$ then denotes all the f values assigned to R elements as set. Of course this doesn't have to mean that f is defined on all R elements. We only collect the values where it exists. So $f(R)$ can even be empty.
We can call $f(R)$ as the image of R by f. In particular, using the full S then the $f(S)$ image means the collection of all values assigned to S elements and this not having all the elements of T means that T is not a subset of $f(S)$. So the $f(S)$ image of S by f can not cover the full T. Then the formal definition is even simpler too: $\forall f\ \neg\, f(S) \supseteq T$.
So the subset relation still helped a lot. This of course includes equality. If however we know that f is only picking up values in T then $\neg\, f(S) \supseteq T$ means actually $f(S) \subset T$.
So with such functions the shortest and also clearest definition is $S < T \,:\, \forall f\ f(S) \subset T$.

2. Among finite sets we have as intuition that not only the part is smaller than the whole, but the above defined f assignments will also remain smaller. Indeed, say we have ten apples. If to seven of them we assign some other fruits say oranges, then these assigned oranges can not become more than seven. They can easily become less than seven if we assign common oranges to different apples. If we don't assign common oranges to different apples, that is our f function is a one to one assignment or so called equivalence then we'll have exactly seven oranges.
Among infinites the part of a set can become the whole with a smart assignment.
The simplest example is the set of the natural numbers. If we leave out 1 then we get only 2 , 3 , 4 , 5 , . . . Now, if we assign to each of these the previous natural, that is f is defined as $f(n) = n - 1$ then the image of this subset that is $f\{\,2, 3, \ldots\,\}$ will become the full set of naturals.

Here again our smart assignment was one to one but we can make easily a non one to one that will do the trick too. For example we could assign to the first two values in our subset, that is to 2 and 3 the common 1 value and then to every n number $n - 2$.
We'll get again all the naturals.
The definition in 1. obviously has only usefulness if there are sets where no smart f can achieve a full image. That such smaller versus bigger infinites do exist was not recognized up until the end of the nineteenth century. Cantor single handedly created Set Theory by not only discovering that there are such smaller versus bigger infinites but at once showing how this fact can prove things that are hard to show otherwise. Indeed, the existence of non fractional, that is irrational numbers was a main concern of already Greek mathematics. Today with our ten based "Arabic" or rather Indian number system extended to infinite decimals and with the added digital method of division process among naturals, we can see at once that the fractions are exactly the periodic decimals and so the irrational are the non periodic decimals. So these not only exist but should be much more.
Cantor's discovery that the decimals are not sequencable, that is to the naturals we can not assign decimals so that all decimals would become assigned, gave a way to see this bigger feeling we get from the non periodic versus periodic. Indeed, the fractions or the periodic decimals are sequencable. Though this itself was not recognized before Cantor either.
To list all the fractions is quite easy by simply going in increasing totals of the numerators and the denominators. With 2 total we have only one fraction 1 as numerator an also as denominator. With 3 total we have two fractions using the numbers 1 and 2. Then with 4 total we can use 1 , 3 or 2 , 2 which gives three possible fractions. So for every total we have only finite many variants and these can all be listed one by one.
By the way, this listing is not one to one. We repeat same fractional values. But this doesn't reduce our result's surprise nor use. Indeed, the fractional values, or only between 0 and 1 could be circled in our list when first encountered and this would be a sub sequence. Cantor's famous proof showed that already the decimals between 0 and 1 are a non sequencable set. This proves that there are decimals that are not fractions without the insight about the periodicalness. But we could doubt the real usefulness of set comparisons from this result alone. The real conquest was to use the same argument for a much later obsession than the Greek irrationality. Namely for the non algebraic numbers.
The algebraic numbers are those that are roots of polynoms with fractional coefficients.
This obviously includes all fractions themselves but lot of irrationals are algebraic too.
So, the idea that maybe all numbers are algebraic is not so far fetched, though was guessed to be false by everybody. But to prove this was only possible by showing concrete infinite decimals that can not be roots of any fractional polynom. This was very difficult.
Yet Cantor's result that the decimals are not sequencable works again because the set of all algebraic numbers are again sequencable. In fact, even the argument is similar as was for the fractions. First we can list all the fractional polynoms and these all have only finite many roots and thus these finite sets after each other give again a list.
So, there shouldn't have been any resistance or hostility against Cantor. Yet there was!

3. The one to one assignment or equivalence is the exact keeper of equality among finite sets. This in itself is a practically useful fact!
Indeed, for huge numbers the actual counting might be much more difficult than a simple comparison. If for example we have a ball room then to decide who are more, the boys or girls, is easy by simply to ask them to form pairs. Which ever is left without pair is more.
For infinite sets things are different because a proper subset of an S set can be equivalent to the whole S. We showed it for the naturals and the subset by leaving out 1.
But we can leave out more and still get a sequence so also an obvious equivalence:

1	2	3	4	5	6	. . . .
↕	↕	↕	↕	↕	↕	
5	6	7	8	9	10	. . . .

Even more surprising is that "half" of the infinite set can be equivalent to the whole. Indeed, the even or odd numbers are both equivalent to the total set of naturals:

1	2	3	4	5	6	. . . .
$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	
1	3	5	7	9	11	. . . .
$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	
2	4	6	8	10	12	. . . .

Amazingly, already Galileo noticed this strange fact when he quite falsely quantized the law of falling bodies. He realized that if a body falls a d distance in a first time interval say second, then in the second it falls 3d, then 5d, and so on for all odd numbers.
This fact is true but: $d + 3d = 4d$, $d + 3d + 5d = 9d$, and so on.
Thus the total fallen distance is always $n^2 d$ up until the end of the n-th second.
This is the real important fact because this way we at once know the total fallen distance up to any t time. Indeed, for example up to **2.7** second the fall is simply $2.7^2 d$.
Recently, I observed a quite shocking consequence of the previous shifting paradox, in fact actually for its original simplest case that is merely shifting the naturals with one step:
Imagine beggars sleeping on the side of an infinite road. Suppose they all have only few coins in their hats in front of them and all have at least one ten cent coin among these.
To steal even this ten cent from one of them would be cruel and probably they'd notice it in the morning anyway. Yet, if we steal that coin from the first beggar but replace it from the second and then replace that from the next and so on, then in the morning we "created" an extra ten cent "out of nothing".

4. A c function on an S set, in other words an assignment of c (s) values to all $s \in S$, is called a choice function on S if for all s it is true that: $c(s) \in s$.
The values themselves could also be called as a sample taken from S if the different S elements are regarded like the full range of the main variants while the elements of these s elements as minor. This is how a chocolate sample box is created for a company.
An actual problem in making a sample set happens if the different s elements have common members. Indeed, we might pick such common element and then these should be regarded twice. But of course sets don't allow duplicates or rather they melt into a single.
With using the c choice function this problem is avoided. Even if $c(s) = c(t) = v$, the c choice function is actually the collection of the (s , v) pairs and so $(s, v) \neq (t, v)$.
The minor problem of how to replace these ordered (,) pairs with actual { } set collections was solved as : $(s, v) = \{ s, \{s, v\}\}$. Ugly but works!
The set of all choice functions on S is called the product set of S and denoted as $\Pi\ S$.
The basic and actually only method of creating bigger infinite sets is a relative claim that uses this product set and also the much simpler combined set concept.
The union of two sets is simply combining all elements of the two sets into one set:
$S \cup T = \{ e ;\ e \in S \text{ or } e \in T \}$
To unite infinite many sets is also easy by first regarding these as elements of an S set.
So actually the s elements of S must be united which means that the elements of these s elements must be combined. In precise terms this means to collect all such t sets of the world that are elements of some s elements of S. To be even more precise, the "some" here meant that "there is", so: $\bigcup S = \{ e ;\ \exists s\ e \in s \in S \}$
The vital theorem using the union and product is called König's Theorem:

If for all s elements of an S set, we know some $k(s) < s$ sets, then the combined set of these smaller sets is still smaller than the product set of S.
So our claim is using the combined elements of the smaller k values which is actually the union set of the k image of S, that is of k (S). So in quite short form: $\bigcup k(S) < \Pi\ S$.

Proof:
We have to show that for any f defined on $\bigcup k(s)$, there is a $c_0 \in \Pi S$ not being a value of f. We'll look at f first only within a $k(s)$.
Since $k(s) < s$ thus, $f(k(s)) < s$ too.
Every $c \in f(k(s))$ picks an element from s, but since $f(k(s)) < s$, they can not pick all elements of s. There is s_0 missing as picked element, and so none of the $f(k(s))$ choice functions can pick this s_0.
Picking such missing s_0 from every s, we get a c_0 that can not be value of $f(k(s))$ for any s and thus, can not be a value of f in the combined $\bigcup k(S)$ either.

5. There is a much simpler, almost trivial relative way to get a bigger set, namely, by simply combining a sequence of $S_1, S_2, \ldots$ sets, if $S_1 < S_2 < \ldots$ stands too.
Then, $S_1 \cup S_2 \cup \ldots > S_n$.
Indeed, an f on S_n can not take up all S_{n+1} elements, by our condition, and thus, can not pick up all combined elements even more.
Interestingly, König's Theorem can create a set even bigger than $S_1 \cup S_2 \cup \ldots$, namely $S_2 \times S_3 \times \ldots$, if by this product we mean the $s_2, s_3, \ldots$ sequences picked from the corresponding sets.
These are mere alternatives to choice functions.
Indeed, a sequence is a function $n \to s_n$ while a choice function is $S_n \to s_n$.
The use of König's Theorem can be done with defining the k function as $S_{n+1} \to S_n$, that is $k(S_{n+1}) = S_n$.
The condition directly gives $k(S_n) < S_n$ and the claim is also directly the claim of König's theorem.

6. The most famous application of König's Theorem is with a "minimal" usage of the condition $k(s) < s$ by $k(s)$ being a single element set, while s being a two element set.
So, we regard any D set and a standard two element set, like $\{ \text{head}, \text{tail} \}$ or $\{1, 0\}$ or $\{ \text{yes}, \text{no} \}$. These fix values can be attached to all d elements of D, that is we regard the $\{ (d, \text{head}), (d, \text{tail}) \}$ pairs of ordered pairs as the s elements.
Then $k(s) = \{d\}$ are the single sets and $\bigcup k(S) = D$ itself.
ΠS is the set of all possible head or tail containing pair choices from the s elements, which actually means head or tail choices for the d elements too.
The yes, no meaning can be regarded as a selection of d elements and so the yes choices as subsets of D.
Thus, we obtain at once that the possible subsets of a D set is a bigger set than D.
This is usually written as $2^D > D$. Indeed, in general, R^D is denoting the set of functions from a D domain to the R range values and so 2^D abbreviates $\{ \text{yes}, \text{no} \}^D$.

7. Using the naturals $\{1, 2, 3, \ldots\}$ as D and the ten digits $\{0, 1, \ldots, 9\}$ as R, the sequences of digits can represent the elements of R^D.
Placing a decimal point in front, we get the possible infinite decimals.
Like: $.3604\ldots$ which also represents a point on the unit interval.
The set of these decimals is bigger than the naturals. In short, we can not sequence all decimals. Cantor's original proof of this went as follows:
Suppose we could list all decimals as:

$$.3604\ldots, \quad .2057\ldots, \quad .0523\ldots, \quad .5379\ldots, \quad \ldots$$

Lets regard the D decimal that has as its n-th digit, exactly the n-th digit of the n-th decimal in the list. So, for our example, D would be: $.3029\ldots$

This D could easily be one of ours in the assumed list.
But now alter every digit in this D, for example, add 1 to all, and thus obtain:
D* = . 4 1 3 0 . . This definitely can not be in our list because it differs from the first in its first digit, differs from the second in its second digit, and so on.
The D number appears as the diagonal number if we write our list of decimals not after each other rather under. Then D* is the anti diagonal or rather "an" anti diagonal, since we had free choices as other digits.
The method itself became known the anti diagonal argument.

8. Cantor's logical question was whether this bigger set of the decimals or the corresponding points of an interval could have a subset, that is smaller, but still not sequencable, that is, bigger than the naturals. In short:
Is there an infinite between the simplest infinity of the naturals and the continuum.
His hunch was the "no" answer, and this became the Continuum Hypothesis.
It turned out to be an undecidable problem, but its mystery remains.
The binary sequences as head , tail outcomes connects to randomness too, and that became a huge field, but it didn't resolve the Continuum Hypothesis.
The identification of the decimals with the actual continuum of the points might already contain some problems. The ancient point paradoxes might have deeper meanings.

9. Cantor explored the continuum that is the interval as point set on its own too.
The famous Cantor Set for example is defined here. So actually Cantor started Topology. Without going into that, we should still mention that the anti diagonal argument can be placed onto the actual line too. So the task is now to see that the points of an interval can not be sequenced. The first point of a sequence in our interval splits the interval into two. Unless we picked as first point one of the end points. This doesn't matter because all we need is a new interval inside that avoids the first point of the sequence. The second point probably is outside of this but might be in or even at the ends. At any rate we can easily find a new interval inside that avoids the point. Continuing this, we get intervals nested in each other and avoiding all the points of our sequence. But these intervals must contain some common points. They diminish in length but might not diminish to arbitrary small length and then we feel that such length of final interval is there too. But even if they do diminish to arbitrary small lengths, a common point must be still there.
As we see a lot of dubious assumptions are present here. To see that not everything is so obvious, we should realize that nested intervals can narrow down to nothing.
This is the case if we use intervals where only one end is included but say the left one isn't. Indeed, regarding the (0 , 1] , (0 , 0.5] , (0 , 0.25] always halved left open intervals, the only common point could be 0 but it is not included in any of them.
Thus Cantor's so called common point axiom for the continuum used closed intervals.
A much better axiomatization was achieved by Dedekind. He assumed that the line is an ordered point set in such a way that if we cut the line into two "halves" then either the left half must have a rightmost point or the right half must have a leftmost point.
Since there are only four possibilities, we could also say that both or none of the halves having last points is impossible. The first would mean two points having nothing between and this is obviously impossible by our intuitions. No end points at all would mean a hole again though less obviously. At any rate, from this Dedekind axiom the Cantor common point axiom follows for closed intervals and so the closedness comes out and doesn't have to be assumed as ad hoc.

10. The precise development of Set Theory relies on the famous Axiom of Choice.
This claims the existence of c choice functions, but as we saw, it is not needed to see the bigger sets, because simple two fix valued functions already lead to this.
The Axiom of Choice, is still vital to get the so called Well Ordering Theorem, needed to compare all sets.

Our definition of $S < T$ was a bit over simplified and could by itself allow quite easily that neither of $S < T$ and $T < S$ stand because we can find functions that image both full sets from the other. This would not be so horrible, after all then we could regard this as their equality in size. But then this should imply the mentioned other original meaning of equal size, the one to one function or equivalence.
The other possibility that both of $S < T$ and $T < S$ stand would be quite contradictory for a comparison. But we couldn't claim either that for any two sets one must exactly stand because of the previously mentioned equivalence or formally $S \sim T$.
So actually exactly one of $S < T$ or $T < S$ or $S \sim T$ should stand always.
To prove this, we better regard a new "flexible smaller" $\prec$ relation that doesn't claim that S is definitely smaller merely that definitely not bigger than T.
And we have an intuitive way to define this as $S \prec T \;:\; S \sim T' \subseteq T$.
Then the claim that $S < T$ or $T < S$ or $S \sim T$ are three exclusive conditions of which one of them is always true, boils down to two claims about $\prec$.
Namely, a fairly easy one that $S \prec T$ and $T \prec S$ implies $S \sim T$ and a very difficult one, that for every S and T at least one of $S \prec T$ or $T \prec S$ must stand.
The final result that the $<$ smaller relation of set sizes is perfect also means that we can then avoid $\prec$ and use instead $\leq$ with the usual meaning that smaller or equal.
The first easy claim is called Bernstein's equivalence theorem, while the second hard one only follows from the mentioned Well Ordering Theorem.
The fundamental idea of this is the following:
We start with some s and t starting elements from S and T that we assign to each other. Then we pick new ones from the outside that is from $S - s$ and $T - t$. We assign these again to each other. We continue this and when we get an infinity of wider and wider already assigned pairs then we simply combine these. Then for such already assigned S' and T' subsets we can again pick from $S - S'$ and $T - T'$.
These widening subsets actually order the elements and so our equivalence is more than an equivalence, it is a $\approx$ similarity between the ordered elements.
The order itself is actually a list or well ordering that has its own rules.
The crucial point is that "eventually" one of the S and T sets must run out of elements and so we actually listed this set fully and part of the other. Thus we created a similarity between one of the full set and a subset of the other that is :
$S \approx T' \subseteq T$ or $T \approx S' \subseteq S$ and these of course imply the same as $\sim$ equivalences.
The crucial problem with this heuristic method is that it requires the newer and newer pickings that is choices as we go ahead. So a timely concept is involved.
The main vision of axiomatic Set Theory is the collection by a property $\{ x \,;\, P(x) \}$.
So, we collect all those x elements that satisfy P. This is indeed spatial and instantaneous.
As it turned out, this collection can become contradictory for some P properties.
The simplest such is $P(x) = \neg(x \in x)$ or in short $x \notin x$. Collecting these x sets that are not elements of themselves leads to contradiction and this is a problem even if we would say that there shouldn't be such sets that are elements of themselves. Indeed, then still collecting these "normal" ones would simply be the set of all sets. So the contradiction is actually in such a full set of all sets but this collection of it as the normals shows the contradiction most directly.
Namely, this $\{x \,;\, x \notin x \}$ set then neither can be normal nor abnormal because one implies the other. Or formally, the basic rule of the property collection $y \in \{x \,;\, P(x)\} \leftrightarrow P(y)$ would lead to contradiction used with our P and y = the normal sets:
$\{x \,;\, x \notin x \} \in \{x \,;\, x \notin x \} \quad \leftrightarrow \quad \{x \,;\, x \notin x \} \notin \{x \,;\, x \notin x \}$
But this doesn't alter the spatial clarity of property collection. In fact, the whole axiomatic Set Theory is simply the systematic restriction of what spatial collections are okay.
Luckily, the amazing result is that we can avoid time here too at the comparison of sets and the step by step widening equivalences can be grasped spatially too!
This is the true technical tour de force of the Well Ordering Theorem.

We have to assume choice functions and usually this is the part that is only chewed upon, avoiding to reveal the deep problems at the time avoidance itself.
Aside from this stupidity, an other typical Formalist feature is that we start with the end.
So regard these orderings on sets on their own and then pull the rabbit out of the hat as: "by the way we can then compare sets too".
I will try to do differently! But first a detour to later results:

11. The infinites as sizes can be regarded as generalized numbers or cardinalities. We already showed that the crucial method of König's Theorem can be regarded as $2^D > D$ with the simplest two element choices. The earlier mentioned and much simpler fact that the fractions are sequencable can also be looked algebraically.
Indeed, first of all they can be regarded as infinite many sequences under each other if instead of our smart method of going through them by increasing totals, we first list all the ones with the fix 1 numerator, then under the ones with 2, and so on:

$$\begin{array}{cccccccccc} \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & . & . & . & . \\ \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \frac{2}{4} & \frac{2}{5} & \frac{2}{6} & . & . & . & . \\ \frac{3}{1} & \frac{3}{2} & \frac{3}{3} & \frac{3}{4} & \frac{3}{5} & \frac{3}{6} & . & . & . & . \\ . & . & . & . & . & . & & & & \end{array}$$

Every stupid decision has its reward and so here too we have the new insight that fixing our stupidity that is sequencing the above infinite table means actually sequencing any sequence of sequences. If the size of a sequence is denoted as ω in general then this would mean that $\omega^2 \sim \omega$.
This is how usually the sequencing of the fractions shown. That is as sequencing an infinite square table. Amazingly, the actual trick can be the same. Indeed we can see the increasing totals as 45 degree increasing partial diagonals of our table. We have to walk through these diagonals after each other and thus encounter all fractions.
But we don't have to use these finite diagonals. We could walk in other manners for example through horizontal and vertical steps and make a turn at the infinite diagonal.

The ω infinity is actually also used as the abbreviation of the sequence as list type.
Then ω^2 should be again a list type namely infinite many sequences after each other.
These types are simply the common name for lists that can be one to one assigned but exactly keeping the order too. We call this being similar and abbreviate it as $S \approx T$. This of course implies $S \sim T$. The crucial and intuitively also obvious fact is that once two sets are well ordered then one has to be similar to a beginning of the other. So among these types we have this as a new smaller or bigger relation that I will mark as $\lhd$.
So, for example $\omega \lhd \omega^2$ simply meaning that ω is a beginning of ω^2.
All this seems so trivial but actually there are surprises! They start with the earlier fact that all beginnings of ω can be cut off or omitted and we are left with the same type.
In reverse too, adding a finite segment to the front will simply melt in: $3 + \omega = \omega$.
Of course 1 , 2 , 3 , . . . 1 , 2 , 3 or in short $\omega + 3 \neq \omega$ because we can not make these similar. This at once shows too that unlike at ω where any beginning is omittable, here at $\omega + 3$ the full ω or $\omega + 1$ or $\omega + 2$ are not omittable beginnings because these would leave only 3 or 2 or 1.

The fully omittable list types are those where all beginnings are omittable and the first such after ω is ω^2 that is, a sequence of sequences. The most obvious fact is that there can not be a maximal omittable beginning of an infinite type because if the B beginning is omittable then B + 1 is also omittable because 1 is omittable from anything infinite.
If we combine all the omittable beginnings of a T then it is either the full T and so T is a fully omittable or the total is still a beginning and then it is the first non omittable A. Cutting this A off we might think that it can not be cut off again from what remained as the end of T but we are wrong. For example in $\omega + \omega + \omega + 3$ we can cut off the same ω three times to alter the list. Luckily, this can happen only finite many times.
Simply because from $A + A + A + \ldots$ we could cut of A. So this shows that in quite general the next fully omittable type after an A is the $A + A + A + \ldots = \omega A$.
After the A, ωA, $\omega(\omega A)$, $\omega(\omega(\omega A))$, . . . fully omittable types the next is quite simply the combined type of these but strangely it is also the same as the sum of these : $A + \omega A + \omega(\omega A) + \omega(\omega(\omega A))$, . . .
If A is the first non omittable beginning of a type then the fact that ωA would be a bigger means that actually the type has to start as $a A$ with an a largest natural number.
Cutting this $a A$ off the end part again has a B first non omittable beginning and it will start as $b B$. What's more, B has to be smaller than A because otherwise $a A$ would have melted into $b B$. Indeed, the last A would be a beginning of the first B and B being fully omittable would swallow up A and step by step all of the A-s.
So every listed or well ordered set is merely an $a A + b B + c C + \ldots$ with decreasing types in this sum. Surprisingly, this also means that these sums have only finite many members. This is a paradox because we just saw that there are infinites and infinites of fully omittable types so we could envision infinite many of them combined except in reverse order. Actually this is the error because we can not envision infinite many in reverse order. Or rather, we have to sharpen our visioning. What helps a little is to realize that already the naturals is an infinity of increasing numbers but we can only list finite many in decreasing order. Once we pick a first, we are screwed and can pick only from a finite many. In a list in general, if we pick an element then we will have infinite many elements before and infinite many after the picked one. So it seems we are not screwed at once. But the very nature of the list that it always requires a next element after any stage of earlier elements, actually means that not just the later ones after a stage must have a first but actually any subset of elements must have a first. Simply because we can regard all stages before encountering any elements of the subset. And then the total of these stages is a maximal such stage and the next element is the minimal of the subset.
A backward going infinite sequence were obviously a subset without minimal element.
Surprisingly though, not having such sequences, at once implies that all subsets must have minimal. Indeed, suppose we had any subset without a minimal. Then we can just pick an element there in the subset. The smaller ones can not have a minimal again because that would be in the whole set too. So we can pick a new arbitrary element again and so on.
Returning to the fact that all list types are $a A + b B + c C + \ldots + z Z$, this gives a method to show that $A^2 = A + A + \ldots A + A + \ldots +$ can be one to one assigned into A itself. And so these types lead to the universal size law of $S^2 \sim S$ for all infinite S.
Remember that this is the sequencability of the fractions when S is simply the naturals.
The simplest consequence of König's Theorem said that $2^S > S$ and this was actually Cantor's first breakthrough again for S as the naturals. The two of course means together that the fractions are less than the infinite decimals so there are irrationals but Cantor used it for also showing that there are non algebraic numbers. The general two laws interrelate in even more amazing ways. Actually this relation of them helps to throw some light on why Cantor's fundamental question whether there is or isn't infinity between the naturals and the continuum is so difficult:

Of course, $2^S > S$ also means that using bigger than 2 base that is as choice values for functions defined on S will be at least as big. But not necessarily bigger because for example the decimals use 10 base and they are the same as 2 because base two binaries can be used instead. So it's pretty obvious that finite bases all remain the same.
The natural guess would be that S^S should be a jump that is $S^S > 2^S$.
That is not true but to prove the opposite that is $S^S \sim 2^S$ would be hapless directly.
The much much stronger yet easily provable fact is that $(2^S)^S \sim 2^S$ too.
Amazingly, the logic is visible for even finite S sets where of course this equivalence is not true but our argument is still valid. This argument is basically: $(2^S)^S \sim 2^{(S^2)} \sim 2^S$.
We'll use S = 10 for demonstration. 2^{10} is simply the number of possible binary ten digit long numbers. So, the ten positions are the S set and the 0 or 1 digits are the choices. Their number is of course 1024. $(2^{10})^{10}$ should be the ten many such possible groups put together, that is the hundred digit long binaries. The number of these is more than the atoms in the universe which shows how among finites this form is indeed a huge jump. Yet in spite of this, the actual number is the same as 2^{100} so indeed the exponent is simply the square. And this combinatorical fact remains among infinites. This proves our claim because regardless that it happens in the exponent $S^2 \sim S$ can be applied. So, the final condensed form of our finding is that: Exponentiation is not sensitive to the B base!
The jump by choices is not dependent on how many choices we have. From 2 to 2^S they will give the same size with S as exponent. This is the only logical possible range for the base because if B is bigger than 2^S then obviously already B^1 gives value above.
Back to the $S^2 \sim S$ law, it of course implies the weaker $\omega S \sim S$ law too.
Now it's time to introduce the concrete infinities after ω.
We actually use the list types themselves as so called cardinals to denote the infinites too. Namely, the first types that have a size. We can not calculate these cardinals by simple type formations and so we just give abbreviated names for them. The first bigger sized list type than ω is denoted as ω_1 and so this the second cardinal after ω. This is already a huge list type much bigger than ω^ω or even using ω power ω many times.
The first next size after ω_1 is ω_2 and so on. After ω , ω_1 , ω_2 , ω_3 , . . . comes ω_ω.
Here we have again that $\omega + \omega_1 + \omega_2 + \omega_3 + \ldots = \omega_\omega$ by simply the melting in.
So this is also a first concrete case of the earlier mentioned obvious fact that for increasing sized sets: $S_1 \cup S_2 \cup \ldots > S_n$.
Here it's even more evident that all the members are smaller than ω_ω and so we tend to jump to the conclusion that then:

$$\omega_\omega = \omega + \omega_1 + \omega_2 + \omega_3 + \ldots < \omega_\omega + \omega_\omega + \omega_\omega + \ldots = \omega\,\omega_\omega = \omega_\omega .$$

A contradiction. But our arithmetical logic was the logic of finite sizes.
Among infinites, definitely increasing the parts will not necessarily increase the total!
On the other hand, König's Theorem does imply that:

$$\omega_\omega = \omega + \omega_1 + \omega_2 + \omega_3 + \ldots < \omega_\omega \times \omega_\omega \times \omega_\omega \times \ldots = (\omega_\omega)^\omega .$$

But this exponentiation is the original size meaning one, that is with choice functions.
This is important because as we just discovered above: $(2^\omega)^\omega \sim 2^\omega$.
Thus $2^\omega \sim \omega_\omega$ is impossible and this is important because the earlier mentioned Continuum Hypothesis undecidability is much deeper. Namely, we can not place this originally guaranteed bigger infinite anywhere in our nice list of cardinalities.
Well the previous result at lest tells something that the continuum 2^ω can not be.
Of course, the Continuum Hypothesis claims that $2^\omega \sim \omega_1$.

Back to the $S^2 \sim S$ law again, S^2 can also mean S many sets combined all having S many elements. So this also means that if we combine an S set as $\bigcup S$ where $S \sim T$ and for all s elements of S we have $s \sim T$ too then $\bigcup S \sim T$ too. From this also follows that if $S \leq T$ and $s \leq T$ for all s elements then $\bigcup S \leq T$ too. So we might simply say that combining can not increase the size. But as we saw above this is exactly a way how we can always get bigger size. The two facts are not contradicting each other.
Best is if we regard the negative version of our first claim:
If $T < \bigcup S$ then $T < S$ or $T < s$ for some s.
A $\bigcup S$ total can be a limit above the elements. Exceeding them all and easily S too.
Such size is called weakly accessible meaning that it is obtainable by simple combining from less many and all smaller sets.
The claim merely says that any size less than such weakly accessible limit will be exceeded by some member or S itself. What does follow from this though is that with increasing totals, that is where $\bigcup S$ is bigger than S and also all s members, there can not be a T size that is exactly the previous to $\bigcup S$. Indeed, S or an s would have to be bigger than T . But above T and under $\bigcup S$ there is no size. Negatively, knowing that an R set has size that has a previous T, means that R can not be obtained as $\bigcup S$ that exceeds S and its members. For cardinals this means that if they have a previous that is they are not limit cardinals then they are automatically not weakly accessible.
So ω_1 that has previous can not be obtained as combining ω many ω sets which is trivial. But our claim is now general and so true for all ω_2 , ω_3 , . . . as well.
ω_ω is of course weakly accessible and needs only ω many members.
We might even jump to the conclusion that all limit cardinals are such but actually this is not a logical necessity. The limit as index of a cardinal is not identical with itself being limit as weakly accessible. It would only be if the index that is the number of cardinals up to it would have to be always less than the cardinal itself. Which seems to be true for all concrete cardinals we created but can not be proved. So we call the hypothetical ones that are so big that their index is themselves as the weakly inaccessible ones.
All this shows that the list types are just seemingly a boring repetitions of going forward one by one. Once we ask the right questions about them then they become alive. The more concrete direction of the right questions is the counter feature of the mentioned restriction that we can only go finite many times backwards.
Indeed, quite to the contrary we can go forward deeper and deeper and sometimes even seemingly rare forward going sub lists can go all the way to the end of the list. These sub list are the so called “cofinals”. The fully omittable list types are just the most primitive special ones. Every beginning being omittable sounds like everything being at the end, but lets call a type itself a cofinal if every cofinal of it must have the same type as the total.
This is a drastically stronger restriction that pushes everything towards the end. Amazingly, it implies at once that such cofinal can not have a beginning that would be the same size as the total. So all cofinals are cardinals. The reverse is not true and indeed, ω_ω is a cardinal that has smaller cofinal namely a simple sequence. The weakly inaccessible cardinals are simply limit cardinals that are cofinals too.

12. The comparability of all sets is the most basic fact of Set Theory and so the clarification of the Well Ordering Theorem is the most important didactical task of mathematics.
As I already explained, the fundamental problem is how to replace the timely choices with a totally spatial set collection. Formally of course we could regard the timely choices as predetermined f and g functions in the S and T sets. So, we have s and t starting elements there and also these functions that assign some outside elements to some S’ and some T’ subsets. By outside we mean from S – S’ and T – T’.
We assign s and t to each other as the start of our h equivalence.
So, our first equivalent subsets are { s } and { t }.

Our f and g functions then pick for these outside elements namely f { s } and g { t }. These are again assigned increasing our h too. Adding these new elements to the start ones, we get the { s , f { s } } and { s , f { s } } sets as second stage equivalent subsets. The next stages would be { s , f { s } , f { s , f { s }}} and { t , g { t } , g { t , g { t }}}. Both have three elements and indeed those are assigned to each other by h.
We get infinite many of these widening sets and then we have to combine them into totals:
{ s , f { s } , f { s , f { s }} , f { s , f { s } , f { s , f { s }}} , . . . } in S and
{ t , g { t } , g { t , g { t }} , g { t , g { t } , g { t , g { t }}} , . . . } in T.
Luckily, this will not cause any problem to establish h because the elements were already ordered to each other. So we can continue the pickings by f and g.
The same combining of widening stages that we did above must always be applied and these combinings have no effect on h. It only grows by the new picked elements.
Having assumed f and g functions that pick the new elements also means that we at once step away from our timely vision a little. After all, we don't choose and those f and g functions are spatially existing choices. To prove that h can come about would mean that we eliminate time completely!
Though using predetermined f and g functions is a big step to avoid the choices step by step it also means that the crucial naïve assumption, we made at our plan is false.
Remember, it was that if we always choose from the outside then since all proper S' and T' subsets have such outside, the pickings could only stop if one of our set runs out that is we reach S' = S or T' = T. Now with f and g given, the running out could happen "any time" if we reach a subset in S or T where f or g is simply not defined.
The amazing surprise is that this seemingly so big problem is not a problem at all!
We can easily find f and g functions that never "stop".
For example, just showing f , let c be a choice function on all the real subsets of S. Real of course just means non empty and so these all have elements and so one can be picked out from each. The fact that the subsets of S are sharing elements is immaterial, because our pickings don't have to be exclusive. Different subsets might have same picked elements. So this c is a totally spatial vision. As an added beauty, this c will not only tell our f but also our s starting element. Namely, s = c (S).
Then if S' is a real and proper subset of S we make: f (S') = c (S – S').
The simple and radical beauty of this breakthrough might falsely make us think that after all this whole affair wasn't so hard. Unfortunately it is still very far from solved.
More strangely, the difficulty of finding the goal that is h from the seemingly so easy f and g is better explained if we refuse the silver platter and so again assume not these f and g just explained from choice functions, rather even more general ones than before.
Indeed, at the start we assumed f and g that pick new elements for some subsets.
Now we start from arbitrary f function, forget S and T and the vision is this:
We fix an s starting element and simply regard how far f can grow from this.
We again don't use f for elements already obtained rather for sets obtained.
So right at the start we regard not whether f is defined for s rather for the { s } set.
If it is, then it still could be that by amazing coincidence f { s } = s.
This would mean that adding the f value to the initial set, that is widening as:
$\{ s \} \cup \{ f\{ s \}\} = \{ s \} + f\{ s \} = \{ s , f\{ s \}\} = \{ s , s \} = \{ s \}$.
So our widening is superficial. We can even envision it as an infinite loop because applying f would give again and again the same. By this heuristic application of any function as a potential widener we ourselves define the natural run or growth of f.
For any given s set and f function, a C set is s , f , union complete if:

1. $\{ s \} \in C$

2. If $S \in C$ and f is defined on S then $S + f(S) \in C$ too.

3. For any $B \subseteq C$ and B non empty, $\bigcup B \in C$ too.

By 3. for the whole C complete set too: $\bigcup C \in C$. This widest element of C is clearly then either such that f is not defined on it or $f(\bigcup C) \in \bigcup C$. Indeed, otherwise, that is if $f(\bigcup C)$ were outside $\bigcup C$ then $\bigcup C + f(\bigcup C)$ were wider than $\bigcup C$ yet by 2. it would be in C. So actually $\bigcup C$ had to be wider that itself.

If we know about f that whenever it is defined its value is outside, then the obvious consequence is that on $\bigcup C$ f has to be undefined. If we also know that f is only undefined on this widest $\bigcup C$ stage then a more surprising consequence is that any element of any S stage in C has to be s or an f value.

Indeed, suppose an arbitrary e element. If it is in all S then of course it in $\{ s \}$ too and so $e = s$. If $e \neq s$ then $\{ s \}$ does not contain e and so the B set of all those stages that avoid e is not empty and so $\bigcup B \in C$. This $\bigcup B$ can not be $\bigcup C$ because then e were not contained in any stage. So, f must be defined on $\bigcup B$ and also take a new outside value. So $\bigcup B + f(\bigcup B)$ is a wider set than $\bigcup B$. Thus it can not avoid e but $\bigcup B$ does and so $e = f(\bigcup B)$.

Our heuristic f defined by a c choice function on an S set was exactly such and an easiest complete C is the set of all subsets of S. And indeed, here every $e \neq s$ element of S is an f value trivially, because $e = c\{ e \} = f(S - e)$.

This example shows that C is not merely the collection of our intuitive growth from s.

Formally this comes about because our rules tell only what must be inside.

In general, any junk could be added to a C and applying 2. and 3. enough times we could get a new complete set.

In truth, our envisioned growth is a set that contains only the sets obtainable by these rules and no junk. At first to find this junk-less minimal seems quite easy.

The solution is the counterpart of union, the so called common part or intersection:

$S \cap T = \{ e ;\ e \in S \text{ and } e \in T \}$ and using this for the C elements of a D set:

$\bigcap D = \{ S ;\ \forall C\ \ C \in D \ \rightarrow\ S \in C \}$

Here I used these “weird” letters because this our application.

D is the set of all possible C complete sets.

Observe that this definition of the minimal complete set is a totally explicit set collection.

We could write the three rules and the common part as a single P property.

Also observe that this $\bigcap D$ common part automatically satisfies the three rules and automatically has to be the smallest such complete set. So indeed we defined the intuitive growth of f from s as a spatial concept.

This of course is still only a subjective conquest. The original goal was to establish a h for any two S and T sets. With our more general vision this means to establish a h equivalence for arbitrary s , f and t , g growths. The mentioned choice function idea to create f and g for two S and T sets then would yield our original goal.

The same idea we used to create the growths can be applied again. That is, start with equivalence complete sets and then their common part would give the minimal.

The big difference is that we have no given h and so we don’t have h completeness.

This h will only come out as the common part.

The new equivalence growth completeness is actually a dual s , f and t , g completeness.

The complete sets are now denoted as W sets of E equivalences:

For these E sets of (a , b) ordered pairs we can use E_1 to denote the set of all of the first members while E_2 to denote the second ones.

1. $\{ (s, t) \} \in W$

2. If $E \in W$ and both f is defined on E_1 and g on E_2 then:
   $P + (f(E_1), g(E_2)) \in E$ too.

3. For any $B \subseteq W$, $\bigcup B \in W$ too.

Again by using 3. for the W set itself, here we can see again indirectly that the $\bigcup W$ widest equivalence stage has to be one where one of f or g were not giving a new pair and so we have equivalence on a widest non continuable stage of the f or g growths.
This might give the following amazing simplification of our whole project:
We don't need the concept of growths at all! Growth complete sets are sufficient!
Unfortunately unlike we had the heuristic choice determined growths and even their trivial complete sets, the set of all subsets, here for the equivalences we don't have any examples of complete sets. So we are back to defining the h equivalence as common part. But it's still valid from our realization that we don't need the growths of the separate f and g.
We can define h explicitly from using all subsets of S and T as trivial complete sets.
This does solve the elimination of time even simpler than we thought and yet we have a major problem. Namely, what if this explicit definition defines an empty set. Since we have no example of any equivalence complete set, this is a very real problem. In fact, we can also realize that our previous subjective success of defining the individual growths was false too. Indeed, just because we have the trivial complete sets as all subsets, it still doesn't mean that a common part exists too.
So the whole common part idea as trivial junk-less set is useless!
It does define what we want but we can not prove its existence yet.
The solution is to relax and yet strengthen the 1. , 2. , 3. rules at the same time.
Relax them so that they do not require completeness in 2. for the widest stages.
This would allow partially complete stage sets that are not continuing on their total.
And such obviously do exist because as simplest cases are $\{\{ s \}\}$ and $\{\{ t \}\}$ as partial f or g growths and $\{\{(s , t)\}\}$ as partial equivalence growth.
Then the really heuristic idea is to combine all these partial stage sets to get the originally aimed minimal complete sets. Of course now the junk can not be avoided any other way than strengthening the rules so that they are blocked out at each partial stage set already.
The good news is two facts: The existence becomes obvious and the combined total of the partial stage sets automatically will be a non continuable stage set by indirect logic if that total is a stage as it should be by 3. So the original 2. rule would stand.
The bad news is that 3. is not inheriting to combined sets so we actually can not use this.
So what we regarded as junk avoiding subjective strengthening before, would now actually become rules that guarantee the inheritance of 3. because we combine widenings that are beginnings of each other.
Unfortunately, for the widening equivalences the set widening is not enough. The equivalence as relation must widen. On the other hand the heuristic choice generated set widenings can be avoided if we work with the equivalence.

13. So we now start from scratch and aim at the equivalence directly to see things clearly:
An R relation is a set of (a , b) pairs.
R_1 denotes the set of the first elements while R_2 the set of the seconds.
A relation is equivalence if every first element has only one pair and vice versa.
A partial equivalence between S and T is simply an E equivalence that $E_1 \subseteq S$ and $E_2 \subseteq T$. If one of them is equality then we call it a maximal partial equivalence between them. It is still partial because a full equivalence between S and T would mean both being equals.
The big claim is that for any two S and T sets we can find an M maximal partial equivalence between them. So, all sets are comparable by equivalence.
The formal solution we'll use is quite simple. We'll find a W set so that:

a). All E elements of W are equivalences so that $E_1 \subseteq S$ and $E_2 \subseteq T$.
b). $\bigcup W \in W$ but $\bigcup W = S$ or $\bigcup W = T$

It's obvious that then this $\bigcup W$ is our M.
Even for someone who doesn't know what $\bigcup$ means but we tell:

$\bigcup W = \{ e \ ; \exists E \ \ e \in E \in W \}$ so these are all those e elements together as set that appear in any E elements of the W. So actually, it is a widest equivalence if the E-s were equivalences and then of course these e are all (a , b) pairs.
What this smart arse plan doesn't tell is how such W set could be obtained.
The letter W at least reveals our intuitive goal.
We regard W as a widening set of partial equivalences. We start with chosen s , t elements from S and T and regard (s , t) as first element in our aimed M maximal equivalence.
Of course (s , t) on its own or rather the { (s , t) } set is a partial E equivalence too.
Then from S – s we pick a new element, from T – t too and assign these again to each other.
This gives a new E again.
As our E increases infinitely we simply combine the elements.
The first elements of an E stage are E_1 in S, the second ones an E_2 in T and so we pick again new elements from $S - E_1$ and $T - E_2$ and they as pair will be added to E.
This plan is first exactified by replacing the step by step choices with predetermined ones:

A c subset choice function on S is simply a function that has as value an arbitrary element for any $S' \subseteq S$. That is: $c(S') \in S'$. Similarly $d(T') \in T'$.
These subset choice functions can tell already the starting elements as:
s = c (S) and t = d (T). So, the initial equivalence is $\{ (c(S) , d(T)) \}$
Then if we have an E partial equivalence as stage, it must be widened with:
$(c (S - E_1) , d (T - E_2))$.
The other vision was that widening stages must be combined.
So if B is such set containing stages only up to a point then $\bigcup B$ is a stage too.
The letter B refers to beginning because it is indeed a beginning of W.
These partial E equivalences or stages melt into the final M maximal.
This simply means that to say that the stage subsets were widened by c , d and union, is impossible to tell in M because we don't know what subsets were these and it wouldn't be true for all subsets of M. The heuristic idea is to do what we did from the start, regard the W set of the stages. Here these are not subsets but elements. Exactly those subsets of M that were stages. So we can talk about them. Now we stepped closer to the formal plan too, that is achieving a W with a) and b).
In fact, our first two intuitive ideas, the start from (c (S) , d (T)) and widenings by new c and d values can be expressed at once:

1. $\{ (c(S) , d(T)) \} \in W$

2. If $E \in W$ and $E_1 \neq S$ and $E_2 \neq T$ then $E + (c (S - E_1) , d (T - E_2)) \in W$ too.

But the third thing we claimed, the combined widenings for B-s is still not expressible!
What are these "already achieved" ones or beginnings of W?
They are obviously subsets of W but how do we tell them apart from any W' subset.
The crucial lucky break coming out of the widening vision is that we don't have to specify them.
Imagine a B up to a point. An arbitrary W' can be two kind in respect of B.
Either W' contains stage that is wider than all stages in our B or not.
In this second case we have again a crucial duality, namely whether B will contain wider stage than all the stages in W'. If again not, then actually W' goes all the way in our B. W' may simply leave out some members in B. So W' is a "cofinal" of B.
Amazingly, then the combined set of W' will still be the same as of B. $\bigcup W' = \bigcup B$.

Indeed, any e element that appears in any element of B will eventually be collected by a wider set in W'. So requiring that the total of these W' is in W means not more than actually the envisioned full beginning sets combined. The other two cases when W' went too far or not far enough, can also be justified by seeing that there are beginnings that go exactly as far as W' does. So:

3. If $W' \subseteq W$ and W' is non empty then $\bigcup W' \in W$.

Now we see at once that 3. applied to W' = W gives that $\bigcup W \in W$.
More amazingly, this at once shows that $(\bigcup W)_1 = S$ or $(\bigcup W)_2 = T$.
Indeed, otherwise $\bigcup W + (c(S-(\bigcup W)_1), d(T-(\bigcup W)_2))$ were wider than $\bigcup W$.
But $\bigcup W$ is the widest in W by definition so actually $\bigcup W$ were wider than itself.
So we instantly achieved b). and we might think that the simpler a). should be easy.
Unfortunately it can not follow from our rules simply because these rules only tell required elements in W but can allow any junk. Our vision actually targeted only those E sets that are achievable by these rules alone containing nothing more.
We could come up with a grand idea to get rid of the junk:
The counter part of combining is the intersection or common part of sets. Just as union, this too can be applied for a whole set of sets: $\bigcap Z = \{ E ;\ \forall W (W \in Z \rightarrow E \in W) \}$.
I used these letters because this is our immediate meaning. Z is the set of all possible junk containing W sets that obey our rules and $\bigcap Z$ will be those stages that are in all W and thus indeed the minimal and avoid the junk.
Unfortunately, there is a second problem with the W-s themselves. We simply can not guarantee that such sets exist. This sounds unbelievable because the first rule is so concrete.
The problem is that 2. on the other hand requires too much. It only allows a non continuation if S or T is reached. But our instinct was good, the starting stage should provide an existence and it does if we relax rule 2. and do not require widening by c and d for the widest stages.
Then of course we don't collect all stages up to M only some beginnings of W.
So our new rules actually try to define the B sets that we were so happy to avoid before.
1. and 3. must merely be repeated with using B instead of W.
But relaxing 2. is not enough because now we have to avoid the junk by rules.
A strengthening of the relaxed 2. would be that for the non widest stage we at once assume the
$E_1 \neq S$ and $E_2 \neq T$ conditions necessary to widen:

2'. If $E \in B$ but $E \neq \bigcup B$ then $E_1 \neq S$, $E_2 \neq T$, $E + (c(S-E_1), d(T-E_2)) \in B$

But this is not enough. We need something strong and what it should be can be found out by returning from the vision to the facts. We'll have to prove a).
We could argue indirectly as follows: Suppose there were some E in B that were not partial equivalences. Then lets regard the first, the narrowest of these and it would have to be achieved by 2'. or 3. But these keep the partial equivalence if the earliers were that.
"This is the first time that this first time happened."
It sounds even funny but what I meant is that up until now we never mentioned that such first or narrowest stages have to be. We didn't need this. And yet this seems plausible because if after every beginnings there is a next then it is the first among the rest. Of course, again just as our heuristic union combining went general, here too we could claim this for any B' subset.
But observe the crucial difference too! For unions $\bigcup B' \in B$ is required while here now we will claim $\bigcap B' \in B'$ so the minimal is inside the B' subset:

4. For any $B' \subseteq B$ and B' non empty, $\bigcap B' \in B'$.

But we still didn't reveal the new grand idea that replaces the failing previous, the intersection.
Now we'll get W as the combined set of the set of all beginnings:
$W = \bigcup \{ B ;$ 1. and 2'. and 3. and 4. $\}$.
So we exorcized time again with a bit longer explicit collection formula but now with an easier provability of its features.
Firstly, the existence is trivial because $\{\{ (c (S) , d (T)) \}\}$ is a B beginning set.
It satisfies 1. , 2'. , 3. , 4.
The second good news about using combinings to get W is that if we can show that all B sets contain only partial equivalences between S and T then W will automatically inherit this and thus obey a).
The third good news is that half of b). is evident for W.
Indeed if $\bigcup W$ were neither S nor T then we could form a new widening exceeding $\bigcup W$.
This now wouldn't contradict rule 2. because it doesn't require widening for the union, but it would contradict that W is the widest beginning.
The other half that is $\bigcup W \in W$ is true because more generally:
For any set of beginnings their combined set inherits rule 3.
4. implies that for any E and F equivalence stages in B the $E \cap F$ common part must be one of them. Indeed we regard $B' = \{ E , F \}$.
This of course means that for any two stages one is a subset of the other.
This at once implies that all equivalences contain the initial $(c (S) , d (T))$ pair.
Indeed, the $\{ (c (S) , d (T)) \}$ stage can not have as subset any other E and so E must have this as subset.