Notes on Elias Zakon's *Basic Concepts of Mathematics*

Even before Volterra commenced his work, the notion that a collection of functions y(x) all defined on some interval be regarded as points of a space had already been suggested. Riemann, in his thesis, spoke of a collection of functions forming a connected closed domain (of points of a space).
Giulio Ascoli (1843-1896) and Cesare Arzela sought to extend to sets of functions Cantor's theory of sets of points, and so regarded functions as points of a space.

Note. Unfortunately, I started taking notes start late, namely, at the section on Dedekind Cuts.

Note. Dedekind's original motivation to develop his cuts was indeed calculus. It was not the continuum, whose lack of existing construction could have felt less uncomfortable than the lack of calculus foundations.

Note. "The first construction of the Real numbers from the Rationals is due to the German mathematician Richard Dedekind (1831 - 1916). He developed the idea first in 1858 though he did not publish it until 1872." [2]
The paper I am partially reading was published later: 1888 "Richard Dedekind (1831–1916) published in 1888 a paper entitled Was sind und was sollen die Zahlen? variously translated as What are numbers and what should they be? or, as Beman did, The Nature of Meaning of Numbers."[1]

Note. Even topology started off before Dedekind published his results, contrary to my guess. First as a theory of dimensions, then with Riemann. But note that "Richard Dedekind, along with Bernhard Riemann was the last research student of Gauss"[2]. The related excerpts from 'History of Topology' are in the Appendix.

The most direct connections are the quotes:

> *The reception and assimilation of Riemann's concept of manifold to the mathematics of the 19th century was slow and inhibited by severe conceptual problems. Of course it was diffi-cult to understand what a manifold in general should be. The easiest way was to translate it as a "number manifold" in the 1870-s and later. At that time the former real quantities had been arithmetically reconstructed by Meray, Cantor, Dedekind, and Weierstrass, and it appeared as perfectly clear to talk about concretely given submanifolds of R'^ or of projective spaces Pfn^ or P^C. Such submanifolds were in the easiest approach defined by inequah-ties as m-dimensional (usually connected) subsets*
>
> *An additional aspect was the problem of compactification of geometrical objects "in the infinite", which in a discussion between Schlafli and Klein was realized, when they debated the difference between one-point compactification of the plane*
>
> *During the last three decades of the 19th century Cantor had devel-oped his theory of point sets in R" in the framework of general set theory. He himself was shocked to realize that bijective maps between real continua of different dimensions can be conceived, and even Dedekind's comforting conviction that more specific maps, in this case bijective and (bi-)continuous ones, would respect the invariance of dimension left the prob-lem to prove (or disprove) such a conjectured invariance. Naive assumptions from space intuition were particularly deceptive in this field; that became even clearer about 1890 when Peano published his example of a "spacefilling" curve with the surprising effect, that the lack of injectivity would even for continuous maps not necessarily lead to a decrease of dimension (or keep it at most invariant), but could as well increase it. Early attempts by Liiroth, Thomae, Netto, and Cantor himself, to prove the invariance of dimension under bijective continuous maps, turned out to contain unclosable gaps and again (as in the case of the continuity proof for uniformization) it was only Brouwer who surmounted the dif-ficulties and indeed proved the correctness of Dedekind's suggestion (Brouwer, 191 la).^^*
>
> *^^ I deliberately use Brouwer's original terminology and do not write R", as Brouwer's terminology leaves the interpretation of the number continuum open. It can be interpreted by classical real numbers, Brouwer's*

> *inituitionistic real continuum, or even (later) by Weylian reals W of 1918.*

Note. This is an important basic distinction and source of constant tension: ‘natural number’ versus ‘natural (geometric) magnitude’

> *We should begin a discussion of real numbers by looking at the concepts of magnitude and number in ancient Greek times. The first of these might refer to the length of a geometrical line while the second concept, namely number, was thought of as composed of units.*[4]

And centuries later, the ‘reals’ emerged, unifying number and quantity and all irrationals (through Stevin and then Wallis)

> *For Wallis there were a variety of ways that one might achieve this approximation, so coming as close as one pleased. He considered approximations by continued fractions, and also approximations by taking successive square roots. This leads into the study of infinite series but without the necessary machinery to prove that these infinite series converged to a limit, he was never going to be able to progress much further in studying real numbers. Real numbers became very much associated with magnitudes. No definition was really thought necessary, and in fact the mathematics was considered the science of magnitudes. Euler, in Complete introduction to algebra (1771) wrote in the introduction:-*
>
> > *Mathematics, in general, is the science of quantity; or, the science which investigates the means of measuring quantity.*
>
> *He also defined the notion of quantity as that which can be continuously increased or diminished and thought of length, area, volume, mass, velocity, time, etc. to be different examples of quantity. All could be measured by real numbers. However, Euler's mathematics itself led to a more abstract idea of quantity, a variable x which need not necessarily take real values. Symbolic mathematics took the notion of quantity too far, and a reassessment of the concept of a real number became more necessary. By the beginning of the nineteenth century a more rigorous approach to mathematics, principally by Cauchy and Bolzano, began to provide the machinery to put the real numbers on a firmer footing.* [4]

Note. Notice the weight of convergence as a crucial element, and a first distant hint about the necessity of a ‘transfinite’ step

> *One further comment by Stevin in L'Arithmetique is worth recording. He noted that, as we stated above, Euclid's Proposition X.2 says that two magnitudes are incommensurable if the Euclidean algorithm does not terminate. Stevin writes about this pointing out what today we would say was the difference between an algorithm and a procedure (or semi-algorithm):-*
>
> > *Although this theorem is valid, nevertheless we cannot recognise by such experience the incommensurability of two given magnitudes. ... even though it were possible for us to subtract by due process several hundred thousand times the smaller magnitude from the larger and continue that for several thousands of years, nevertheless if the two given numbers were incommensurable one would labour eternally, always ignorant of what could still happen in the end. This manner of cognition is therefore not legitimate, but rather an impossible position ...*
>
> *Further progress in the development of the real numbers only became possible after ideas of convergence were put on a firm basis. However, there was a strong influence in the other direction too, since progress in rigorous analysis required a deeper understanding of the real*

> *numbers.*[4]

Note. Origins of the completeness axiom, and also notice the distinction between completeness: the existence of inf/sup and convergence: the sequence converging to inf/sup (which is different from the existence, since a bounded seq could have a sup but not converge to it .... e.g: alternate sqrt(2) approximations with negative random numbers.)

> *... though Cauchy implicitly assumed several forms of the completeness axiom for the real numbers, he did not fully understand the nature of completeness or the related topological properties of sets of real numbers or of points in space. ... Cauchy did not have explicit formulations for the completeness of the real numbers. Among the forms of the completeness property he implicitly assumed are that a bounded monotone sequence converges to a limit and that the Cauchy criterion is a sufficient condition for the convergence of a series. Though Cauchy understood that a real number could be obtained as the limit of rationals, he did not develop his insight into a definition of real numbers or a detailed description of the properties of real numbers.* [4]

Note. It is crucial to understand the following quote exactly, noting the inter-relations between convergence, limit, existence of a limit, properties of real numbers, definition and construction of real numbers.

> *Cauchy, in Cours d'analyse (1821), did not worry too much about the definition of the real numbers. He does say that a real number is the limit of a sequence of rational numbers but he is assuming here that the real numbers are known. Certainly this is not considered by Cauchy to be a definition of a real number, rather it is simply a statement of what he considers an "obvious" property. He says nothing about the need for the sequence to be what we call today a Cauchy sequence and this is necessary if one is to define convergence of a sequence without assuming the existence of its limit.* [4]

Note. Is it insightful to check what exactly was incorrect in Bolzano's original unpublished theory?

> *Bolzano, on the other hand, showed that bounded Cauchy sequence of real numbers had a least upper bound in 1817. He later worked out his own theory of real numbers which he did not publish. This was a quite remarkable achievement and it is only comparatively recently that we have understood exactly what he did achieve. His definition of a real number was made in terms of convergent sequences of rational numbers and is explained in [22] where Rychlik describes it as "not quite correct". In [28] van Rootselaar disagrees saying that "Bolzano's elaboration is quite incorrect". However in J Berg's edition of Bolzano's Reine Zahlenlehre which was published in 1976, Berg points out that Bolzano had discovered the difficulties himself and Berg found notes by Bolzano which proposed amendments to his theory which make it completely correct. As Bolzano's contributions were unpublished they had little influence in the development of the theory of the real numbers.* [4]

Reminder. Keep on the lookout and remember this quote once we learn the difference between convergence and uniform convergence. My guess is that uniform convergence is the 'natural' 'obvious' kind, like when approximating the sqrt(2), whereas convergence is a but wilder? jumping around about the limit but still converging? I am probably wrong, but it is good to have a theory to refute later but hook onto for now.

> *Cauchy himself does not seem to have understood the significance of his own "Cauchy sequence" criterion for defining the real numbers. Nor did his immediate successors. It was Weierstrass, Heine, Méray, Cantor and Dedekind who, after convergence and uniform convergence were better understood, were able to give rigorous definitions of the real numbers.* [4]

Note. The last part of [4] is entitled 'Attempts to understand'. This reminds us to keep in mind that the standard definitions work excellently, but this does not mean we understand them in terms of mathematical logic (consistency doubts included) or even informally. Epple summarizes the problem:

> *What was a real number at the end of the* 19th *century? An intuitive, geometrical or physical quantity, or a ratio of such quantities? An aggregate of things identical in thought? A creation of the human mind? An arbitrary sign subjected to certain rules? A purely logical concept? Nobody was able to decide this with certainty. Only one thing was beyond doubt: there was no consensus of any kind.* [4]

It is interesting to note the parallels with physical 'models' of the universe, which suffer from a similar problem. The examples are countless, I can at least remember two. Newton was criticized for not 'explaining' gravity and its action at a distance. Einstein postulates in 'On the Electro-Dynamics of Moving Bodies' was subjected to similar objections initially.

> *The objection that Einstein did not give a truly physical explanation of the relativistic effects, but transformed away the problem by a maze of physically unmotivated mathematical formulae, was removed a few years later, when H. Minkowski (in 1908-09) discovered the deeper significance of the two postulates of Einstein.* [5]

Note. Although we are not directly concerned with the 'Attempts to understand' in this document. It is useful to remember the following construction as a memory peg for that chapter.

> *We'll construct a certain real number which, although historically not one that was looked at, will let us understand some of the questions that arose. Let us start with the 100 two digit numbers. A simple code will let us translate these into letters, 00 become a, 01 become b, ... , 25 becomes z, 26 becomes A, 27 becomes B, ... , 51 becomes Z, then code all the punctuation marks, and then make all the remaining numbers up to 99 translate to an empty space. Now create a number, say c, starting from the 100 2-blocks.*
>
> $$c = 0.010203040506070809101112131415161718192021222324 25...$$
>
> *Then continue with the 10000 pairs of 2-blocks 0000, 0001, 0002, ..., 0099, 0100, 0101, ...*
> *Then the 1000000 triples of 2-blocks etc. We can represent c as a point on a line segment of*

> *length 1. Yet every English sentence ever written or ever to be written, occurs in the decoding of c into letters. For example "one third" has 9 characters so will be decoded from c around 1018 digits after the decimal point. This article is there, both with the misprints which inevitably occur and a corrected version is there (but one has to go rather a long way to the right of the decimal point to find it!). The whole of Shakespeare is there, as is every book yet to be written, etc!* [4]

Some paradoxes and their resolutions are mentioned in the Appendix.

Note. One of the latent intuitive difficulties that bother the mind is the known fact (an application to the completeness axiom in Zakon's book) between any two rationals there exists an irrational and vice versa. This causes no discomfort to the mind until one remembers that the rationals are countable while the irrationals are not. This contradicts our finite experience, since alternating a finite number of apples and sticks requires that the total number of sticks and apples have a difference of exactly one. To lessen the discomfort, we should remember, as we deduced in 'On the Necessity of Infinity', that all transfinite steps are cherry picked extrapolations of the finite experience. It just happens that the chosen extrapolation of one-to-one correspondence as a basis for cardinality, disallows taking our intuition about the finite betweenness with us to the transfinite.
Furthermore, we should keep in mind that *uncountable* really means *unindexable*. For reasons of language, this reduces the discomfort even further when relating to the stick and apple analogy. In fact, this replacement almost destroys the analogy. For we are counting -and not merely indexing, but measuring a totality, albeit finite,- in the stick and apple case, whereas in the rational and irrational case, we are comparing two infinite sets with regards to indexing by the natural numbers. Both sets being infinite, comparison by counting a totality is senseless, and of course, this brings back the one-to-one correspondence extrapolation (to the rescue).
To explain the term *unindexable* even further, it is enough to remember the common countability (indexability) proof of the rational numbers. When such an indexing map cannot exist, we simply declare a set an unindexable. The obvious example is the diagonal argument and we guess that in all other such proofs it will be necessary to take the *transfinite step*.
Coming back to the apple and stick example, we have to notice that it is actually impossible to exactly achieve our goal; there will always be one extra apple or stick. This has two consequences: it weakens the analogy even further, and more importantly, this could be seen as another fundamental difference between finite and infinite sets. In fact, this seems very related to another usual definition of an infinite set: a set which supports a non-surjective injection into itself (see problem 1.§9.4-5 in the book).
Finally, Cantor was German, so uncountable is merely an agreed upon translation. The terms Cantor used were *überabzählbar* for uncountable and *abzählbar* for countable. This fact actually reduces the discomfort even further because a literal -albeit rough- translation could actually be over-denumerable and denumberable, and that is not far from our alternative words.

Note. Once one understands the following three things:

- The amazing power of set theory comprehension (set-builder notation), especially the granted existence of any set one can propositionally define in a finite number of words.

- The essence of the transfinite step.
- The countability proof of the rationals (the easiest on the list).

The diagonal argument falls out much more naturally.

Note. The trichotomy of rational numbers was first explicitly showed by Dedekind [7]. This provides a hint to the Dedekind cuts, remarking their closeness to trichotomy.

Note(2). Even after I fully understood the idea behind Dedekind Cuts, and why simply taking the set of all supremums and infemums of upper and lower bounded sequences is not as elegant since it leads to the necessity of using equivalence classes, I still hanged. Proving that the set of all cuts in a complete ordered field is, as Zakon puts it, a drudgery indeed. I had started proving this without help from the book, and did this for addition. But once again, I was raw. I used the definitions directly, and also defined addition in the most direct way possible. Adding a cut A,B to a cut A',B' involved set A''=A+A' and B''=B+B' elementwise. the proof was long. Also, I still was quite uncomfortable with Zakon using supremums as a tool to start doing all the necessary proofs since, I thought, he was unwarranted to compare rationals to gaps. I thought hard about this when I regathered the spirits to try to prove again that the set of cuts is what it should. I did see that the construction is very subtly different from the idea of a supremum (or infemum). But still technically, I found it wrong to declare the 'non-existing' gaps as supremums. Something was wrong. At the same time, I should have noticed that I would be using sup/inf like proofs all the time in the proof at hand. Had I been creative, I would have tried to construct a structure around the cuts that would allow the usage of supremums in the proofs, and that is exactly what Zakon did, but the idea is so subtle that it escaped me, even after several readings. One could do away with this, and proceed with the 'raw' proof like I did, but this would generate many more pages and a lot more work. But creating a sup/inf device from the beginning is simply too elegant. And I decided to retry by first trying to create such a device on my own, if at least to lessen the amount of toil ahead. To be fair, had I come up with this idea on my own, I would have been impressed. Nevertheless, this is yet another sign of lack of creativity on my side, and a good lesson. One could blame Zakon for not motivating his device properly, but he hardly ever does. This causes struggle, but I like to think it is of the good kind, since I hope it improves my absorption of the material.

Note(3). None inductive 'infinity'
Q: In the solution of the following problem, what is the precise difference between what is requested and what a 'proof' by induction would furnish.

Problem 2.§9.11. Prove the *principle of nested intervals*: If $[a_n, b_n]$ are closed intervals in a complete field F, with

$$[a_n, b_n] \supseteq [a_{n+1}, b_{n+1}], n = 1,2,\ldots$$

then

$$\cap_{n=1}^{\infty} [a_n, b_n] \neq 0.$$

Problem 2.§10.4. In Problem §9.11 show that if the intervals $[a_n, b_n]$ also satisfy (for fixed $d > 0$)

$$b_n - a_n \leq \frac{d}{n}, n = 1, 2, \ldots$$

then

$$\cap^{\infty}_{n=1} [a_n, b_n] \text{ contains } \textit{only one} \text{ point, p,}$$

and this p is both $\sup a_n$ and $\inf b_n$. Also show that, if F is only *Archimedean*, the same result follows, *provided that*

$$\cap^{\infty}_{n=1} [a_n, b_n] \neq 0.$$

TODO, rewrite the whole answer.

1. no founded infty symbol before foundations. Therefore no real proofs using the concept, or equivalent ones such as the continuum.
    a. divide math into
        i. what does not involve such concepts in proofs (using the 'real number line' included?).
        ii. math that relies on such concept but before they were founded (calculus).
        iii. foundations, refounded calculus, and any other mathematics that sprang out of these foundations. Do I have an example?
    b.
2. remember goal: calculus
3. extrapolation device, here we are thinking about the formal proposed system (what did we just do here?) this is logically not necessary
4. some infinite processes can be described in finitely many words / symbols
5. naturals and induction
6. what is not inductive
7. procedural method: see inf symbol, not use induction to prove, disproving is ok
8. behind the procedural method, transfinite thinking, try to actually totally justify it in terms of contradiction
9. equivalence of nest. inter. with completeness axiom
10. ultimately, reduce to thinking about completeness, a good robust ground
11. develop higher concepts, build it in, we do not want to have to reduce every time
12. Wittgenstein's Sign-games?
13. specifically to the problem, the number of intervals must be countable!
14. The difference between sqrt(2) and all of its approximations (each of them having a finite number of digits, but the fact that there is infinitely many of them (converging)) "He also defined the notion of quantity as that which can be continuously increased or diminished and thought of length, area, volume, mass, velocity, time, etc. to be different examples of quantity. All could be measured by real numbers." note that by the density of rationals, already treating numbers individually gets us into trouble 'moving', we need intervals? Make this idea less vague.
15.
16.

A: Applying induction to 9.11 trivially establishes that $\forall n \cap^{n}_{i=1} [a_i, b_i] = [a_n, b_n]$. To see that this

result not imply what is to be proved, we have to finesse the difference between writing '*for each and every n'*, and '*when n is infinite'*. While it is tempting to also consider '*as n tends to infinity'* there is really nothing that warrants the use of '*tends to*' in this context. The two expressions, while at the shallowest depth of analysis seemingly equivalent, differ in exactly the same way in which an ordered field and a complete ordered field differ; the way in which the rationals, indefinitely divisible and dense, can nevertheless only ever approximate the square root of two. Under closer examination, the expression '*when n is infinite'* is misleading indeed. n is a natural number, and *infinity* is not. The set of natural numbers is (already by definition if one wished) inductive, while the first and foremost characteristic of transfinite numbers under which our *infinity* (or more specifically $\aleph_0$) falls is that they are not. There is a crucial mental step here that *must* be taken: the step to the transfinite, the step from a never-complete *process* to an imagined *completion*, the step from *potential infinity* to *completed infinity*. ???This step need not be uncomfortable or mystical, we can reason about it in a finite number of sentences, and to make it even less ??? we can ... not wild ... but some 'rule', we can talk about the rule about the process even if executing the process would never end ... ???
Imagine the completed process of taking all the intersections in question. Looking at our inductive result, we find it impotent. There is no connection between any number n and its intersection interval in the induction (any of the natural numbers *for each* we have established a resulting intersection interval), and the trans-finitely imagined *last* interval, in just the same vein, there is no least positive rational number of the form $\frac{1}{n}$.
To be able to hold on to??? '*when n is infinite'* , we *must* concede that n is not a number. Hence, any of the number *for each* of which we have established a result, is not related.

Note. Spivak proves the Isomorphism of all complete ordered fields in a much shorter way than Zakon. I had decided this would be the only chapter I would not do the exercises for in the Zakon book, because it felt it to be too de-motivating (a feeling that I very seldom act upon).

Note. I have finally aligned myself with Zakon's way for the (optional) Dedekind Cuts section. From the experience I gained in this long struggle, I once again noticed the importance of 'de-tersifying' Zakon. Now that it all falls into place, I can spot the important sentences. But Zakon does not try to emphasize them, or colorfully describe them as being in any way more important than others. This might be a good thing since such a judgement might be subjective, but also as I already mentioned in Note(2), it makes the struggle worthwhile. Another thing that Zakon does not do is providing an outline, a plan of attack, making sections like this one feel very serial. This again needs 'de-tersification' on my part and is a great way to prepare for even terser texts and material.
Getting back to the section's topic, I was able to realign myself by using the method of trying it myself and comparing with the book when I was happy with what I did. Taking the hints mentioned in Note(2), I proved the complete ordered *set* result. My construction was almost identical to Zakon's, but less elegant. The cut I propositionally constructed to be the lup of an upper bounded set of cuts was identical to his. But I defined it directly, my definition being more complicated, whereas Zakon split it into multiple insightful notes. Here is where outlining his goal would have helped (Spiral reading is here to stay). I am slowly learning to wield the power of predcative set constructions. But using paintings as an analogy, whatever I come up with

ends up looking like a five year old's stick figure compared to a master's portrait. Nevertheless, I feel progress. I had decided to treat Zakon's set R-bar as simply the set of all cuts. His justifications of 'old' and 'new' elements still feeling to me like a shortcut not to have to define an isomorphism between the rationals and their cuts in R-bar. In any case, I constructed the lup cut having the 'lower' set as 'the set of all rationals x such that x<=y for all rationals y belonging to all the 'lower' sets of all the cuts in the bounded set'. This is indeed Zakon's construction, but in a much more direct and 'childish' form. I courageously defined the upper set as the set of all upper bounds of the lower set, using the full freedom of propositional set definitions. Zakon's construction was exactly the same, his proof of completeness in essence also the same.
Having done this I proceed to work on addition of cuts by myself, the way I defined it and proved was the way Zakon did which, after re-reading parts of the section again, made me realize that I am now aligned with the section. Again, I really like Zakon's approach, even though it took longer to appreciate this time. Alignment achieved, I decided to drop my plan of simply proving the the resulting set is a complete ordered field on my own, but from here on following the book again, first rederiving all the section's field proofs, and then the remaining proofs (the ones that are left as exercises). I was indeed wondering what kinds of exercises this section might have except for proving the axioms, and I was happily surprised to be right.
I had to go to Dedekind's paper to compare with Zakon's. Dedekind used a slightly different construction in terms of the intersection of the cuts, resulting in the rational being either in the lower or upper cut, simply by choice. I also read in Spivak's yet another slightly different construction. I see the freedom one has to use the cuts to produce the reals. Per example, I envisioned a variation on Zakon's and Dedekind's. In Zakon's the 'new'/'old' issue is there and it is justified in more length than usual and in 'english'. In Dedekind's the choice mentioned is an issue. What about a definition of the cut similar to Zakon's but using 'strict' lower and upper bounds (so < and not <=). This would result in a rational producing a gap as well. All cuts would be 'gaps', and this would unify the proof (e.g: addition of gap to non-gap) I hoped. In this case the union of both cut's sets in case of a rational, is no longer the set of rationals since the rational itself must then be 'added' to complete the set of rationals. In case of an irrational the union would still be the complete set. Maybe this could be seen as characterizing the cut (an alternative definition of 'gap'). Using this, I found the proof for addition not much improved, in fact, the proof was very slightly more verbose since when a rational is not in the lower set, it now can either be in the upper set, or be the cut's rational itself (if it is a rational cut).
Nevertheless I found it a nice idea. Zakon's gaps made me go back to Dedekind's paper and search for 'gaps', Dedekind surely did not define them in exactly this way, and he mentions them very little, only as 'gaps' in the straight line.
Originally I thought that it is the sup/inf idea of Zakon that will shorten the proofs of this section compared to my original attempt (see Note(2)). But Dedekind also started with an outline for cut addition, did not use inf/sup but still produced an extremely short proof. Unlike me and Zakon, he defined addition in yet another way, using <= to make sure the lower set 'by construction' contained all lower bounds of the upper set. My definition did not do this, and hence I had to prove that the lower set indeed captured 'all' the lower bounds, in the process of proving that the two defined sets constitute a cut. Here was an excellent example of three definitions, from crude to elegant to 'over-produced'. One does need some long sightedness, some prediction of the results of his definition, or, as an alternative, one must try the definition, find out how it can be

improved and spiral on it. In any case, the power of propositional set definitions is for me still like an Excalibur sword held by an unworthy. I hope being aware of this helps. From this point of view Zakon's split of what should be an exercise and what not in this section makes perfect sense and shows great textbook writing skill, understanding the (low) level of students.
As a last note, I noticed that Dedekind defined 'chains' and according to [1] this is one of the most important contributions of the paper, however I never heard of them (they are at a more fundamental level than real analysis: logic). Still it is amazing how evolved Dedekind's paper is and how simplified it reaches a first real analysis textbook. One would think one knows what Dedekind did. But one is miles away, merely using one distorted and polished screw from Dedekind's large construction. I wonder though in what context is that larger construction needed outside of logical foundations of real analysis, and how it relates to him writing:

> *The majority of my readers will be very disappointed in learning that by this commonplace remark the secret of continuity is to be revealed*. [8]

Is this a case of expanding a commonplace idea for the sake of a seemingly monumental work? (I am surely wrong).

Note. I found a thesis about programming the Dedekind Cut construction of the reals in Coq (http://www.ps.uni-saarland.de/~hornung/bachelor.php). Is that not an excellent project to round this book, while continuing with the History of Analysis book before jumping to Zakon's second book? I wonder what one can 'do' with this construction in Coq, just like I wonder about the practical value of the construction in general sometimes, except as a proof of existence and for peace of mind. After all, mathematics did well without out for a very long time. And it is still quite useless to all things computing. Or is it not? Floating point, what a joke of a number model! But the reality for computing. I do think about the contrast at times and if it will change.

Note. It it good to note and remember, how elegantly and nicely Dedekind cuts can be proved to be complete. This happens immediately after defining the ordering relation, concluding they are a complete ordered set, and before passing to the field axioms. The proof boils down our axiomatized completeness to easily constructing for a bounded set of cuts a cut that can be proven the be bounding it, and also, proven to be the least by simple contradiction, showing that any lesser bound would fail to be an upper bound. This is so beautiful that for a second, it makes one forget that there is here a 'transfinite step' involved. Indeed, where did it go? Again using the power of propositional set definitions, it is in the construction of the cut set that the step is now 'hidden', since the cut's lower set is the set of all elements less or equal to any element in any lower set in the sequence, which is infinite. Hence, our construction only makes sense as a completed infinity. I think it is crucial to remember this, including the fact that defined this way it feels much less objectionable: an illusion due to the 'hiding effect'.

A simple way of seeing the effect is also that a cut by itself, is an infinite totality, reduced to one 'completed' object. But herein lies one of the very important ideas: that it is an infinite totality that we can always finitely describe, as compared to an 'unstructured one'. This is indeed the base of the work-in-progress (probably for ever) Note(3). Indeed, the idea is in a way a triumph on the level of language precision, maybe more so than on the level of mathematics.

Note(2a). Zakon indeed created a perfect system for his construction.
The set Rbar is the set of 'old' and 'new' elements. At the same time (and this is brilliant as we will explain shortly) it is the set of all cuts in R.
The 'old' elements are the ones determined by rationals, the 'new' ones are the ones determined by gaps. 'old' elements (e.g: a) determine the cut a det. ({x<=a},{a>=a}). It is trivial to show that a rational determines one and only one cut. This is helped by the intersection properties of the cut lower and upper sets, an intersection if not empty consists of exactly one element, and that is then the determining rational.
There is a one-to-one correspondence between the elements in Rbar ('old' and 'new') and cuts. This is true because each 'old' or 'new' determines a cut (as shown for old and by definition for new), and because each cut either is a gap (then it is determined by a 'new' element) or it is not, in which case it is determined by the rational.
This is subtly powerful since then Zakon proves that a cut is the sup/inf of a set of rationals. in the proof, he of course uses a comparison between a rational's cut and another cut to 'speak for' the relation between the rational cut's rational in R and the cut in Rbar. This is a crucial element that removes ambiguity in all later proofs, bridging cut-like set of rationals and elements in Rbar thought a sup/inf with total precision.
Additionally, any such sup/inf is proven to exist by the completeness of Rbar (more on that later). This is another brilliant idea since in proofs we can start with rationals or gaps, form sets of rationals, (we use that for all the definitions like addition, inverse, etc.) and the sup/inf is guaranteed to exist, and it is a cut, which can then be used in the rest of the proof, irrespective of the cut being a gap or not (no cases needed).
About the completeness proof: This is yet another brilliant subtlety. Any bounded cut, is linked to rationals by first showing that if it is bounded, it has bounds that are 'old' since we can construct a cut by taking any rational from the upper set (brilliant, and might even be the essence of inter-density of rationals and irrationals). These bounds can all be grouped into an upper and lower set of yet another cut. It is then trivial to show that this cut is a lup, by considering a lower lup and showing it leads to a contradiction. All this is so elegant that it has taken me so long to get it.
Spivak correctly points out

> *This means, in particular, that a rational number (a member of Q) is not a real number; instead every rational number x has a natural counterpart which is a real number, namely, {y in Q: y < x}. After completing the construction of the real numbers, we can mentally throw away the elements of Q and agree that Q will henceforth denote these special set. For the moment, however, it will be necessary to work at the same time with rational numbers, real numbers (sets of rational numbers) and even sets of real numbers (sets of sets of rational numbers). Some confusion is perhaps inevitable, but proper notation should keep this to a minimum. [9]*

Zakon would have never motivated this section with an explanation such a this one. This shows how important it is to have support books.

Note. When thinking of the completeness proof for Dedekind cuts, it is useful to keep in mind how, for the rationals, $\surd 2$ is used as an example of non-completeness. It then becomes apparent how, in a sense, the cuts construction remove a certain ‘restriction’ inherent in the definition of rationals.

Note(2b). Concerning the completeness proof in Note(2a). It is interesting to find Dedekind’s original equivalent from [8]. As Feferman points out,

> *The completeness axiom is equivalent to the Dedekind continuity condition, that there are no gaps in the line.* [11]

and Dedekind’s continuity condition appears at the onset of his paper where he says:

> *I find the essence of continuity ... in the following principle:*
> *‘If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions.’ [8]*

I had read this, but I failed to spot the connection. This appeared to me as merely the definition of a cut. Undoubtedly though, seeing the difference is important and extremely subtle. The upper class in Dedekind’s continuity condition is exactly the set of all upper bound rationals in the completeness proof. His ‘one and only one point’ is the trivial supremum condition. I now additionally notice that his ‘point’ seems to be geometrical. I am lead to this thought by Feferman’s remark

> *To sum up, from the point of view of conceptual structuralism and just as with Cantor’s construction, Dedekind’s construction and characterization of the continuum is not a basic conception but a hybrid of geometric, arithmetic and set-theoretical notions. [11]*

In Zakon’s book, this ‘point’ would be a cut, either ‘old’ or ‘new’, which makes sense since it is the geometrical point that we were going after all this time (informally).

Note. I previously wondered whether Cantor’s transfinite numbers are ‘scientific’, and also if, in practice we ever need anything other than rationals. Without getting into the details, here is an interesting related quote from Feferman:

> *As a result of such work, I proposed in (Feferman 1988, 1993) the following Conjecture. All (almost all) scientifically applicable analysis can be carried out in W.*
> *Of course, that is by no means all of analysis. There is a lot of analysis which goes beyond any potential scientific application, including analysis on non-separable spaces and analysis which involves non-measurable sets, neither of which can be done in W. So one must consider test cases in the applications of analysis to science in which those kinds of things might figure. For example, Itamar Pitowsky has proposed a use of none-measurable sets in quantum mechanics, but there is considerable dispute as to whether that is a reasonable model. Possible uses of non-separable spaces in other aspects of quantum theory have been suggested by G.G. Ench - and again those are very speculative models, which are by no means generally accepted. Otherwize my working conjecture is at least corroborated in settled scientific applications of analysis. [11]*

The ‘W’ he refers to is:

> *In particular, I have developed such systems in which you can redevelop substantial portions of classical analysis and modern analysis on predicatively justified grounds. One such is a system I call, in honor of Weyl, which is in a certain sense a variable finite type extension of ACA0. [11]*

Note. Considering completeness from the opposite point of view, we can first think of

Dedekind's continuity in Note(2b), and the geometric intuition of a bounded set having necessarily a sup (the 'point' beyond which we cannot move any further) as Zakon presents it, we find that it is indeed natural to think that numbers would be able to 'represent' this idea. Forgetting all 'transfinite steps', and in Brouwerian/Wittgenstein style (as far as I can remember) considering this continuity to be an essential intuition just like (and different from) counting. It then becomes surprising that $\sqrt{2}$ is not rational, instead of being surprised about the need for an esoteric *transfinite step*; the defect is then seen as being in the rationals which becomes themselves the 'black sheep'. A new explanation suggests itself. We must not separate 'number' from operation, and must always consider them together. The operations on rationals support the existence of finitely describable processes that never terminate ($\sqrt{2}$). Firstly, This leads us directly to the topic of termination of algorithms. Secondly, classical analysis then starts to look like an achievement concerning the precise description and ability to comfortably work with any kind of 'algorithm', getting us out of the problem we put ourselves into with our too simple number system that we based on arithmetic with naturals, unsuspecting the (inevitable) invitation of 'real' infinity, as compared to the 'simple' infinity of the naturals and rationals (elucidated as being the same by Cantor). This is achieved by baking the algorithm into the predicate of the cut, with the requirements that it is humanly describable and provably cut producing.

Note. I found yet another internal block. It turns out I was still thinking of mathematics with calculations in mind. But I realized that real analysis, and most mathematics for that matter, should not be thought of in this way. The thought would always bring me back to try to relate real analysis to 'calculations' which is a good thing, but not at the onset of the meta world's discovery. Only later in the process, once the meta-world is known, should one see what fruits one can reap, because otherwise, progress in the meta-world is always half-hearted and not the best it can be. The whole merit of these meta topics is to 'zoom out', elevate the mind, over 'calculations'. From this point of view, real analysis is meta-calculations and the real numbers are meta-numbers. It allows 'knowing' everything there is to know about 'calculation numbers' (by the way of correspondences), and a lot more. And keeping that in mind is important. An analogy is considering the world of calculations (even with variables, so classical algebra and transcendental functions, etc...) as the surface of a sphere. In this world, an infinite number of roads, impasses, etc... exist. But for somebody living flatly in it, they can never see very far. Real analysis, instead of spending all the effort charting the world by trying to discover in on that surface, first elevates itself (using the completeness axiom per example). This elevates sufficiently high for a bird's eye view. With real analysis, we can say things about calculations that we would never be able to say by only doing calculations. It is hence on a meta level. Think of trying to do practical calculations with Dedekind cuts. It is surely quite impossible. On the other hand, try to analyze the 'real number field', 'continuity', 'calculus' using calculations... We must be thankful for this beautiful discipline, and we can only be so once we understand that the label 'real analysis' has 'meta' in it.

At any time that we go to a meta level, we enter a new world, created by us, and at the beginning not very familiar (higher dimensional spaces, variational calculus which now I see as a meta topic of calculus itself, …). It is an unknown world, and we should think of it as such

when we explore it. In fully controlled schizophrenia, we leave the practical world, the applied world, the meta topic's origin world, and we try to fully understand it, fully describe be most elegantly, fully find and solve the world's own specific special meta (abstract) problems. However, we would be totally in the dark, and be able to make extremely little progress if we deliver ourselves to its pure abstract nature. Therefore we rightfully still use the original world's ideas as a white cane, to do some progress initially, leaving a secret thin thread connecting the two worlds. Eventually, we know the meta world well enough, and we can use in the most wonderful ways.
It is very strange and weak of me that this thought (finally) occurred to me in this clear form only now.
This kind of attitude, even when taken, does not clear all challenges studying Zakon's book. The analysis by history book is of great help, but synchronizing both is something that is not easy to manage, and the biggest challenge occurs when, despite being in the 'meta' personality, a concept or a section is introduced, which, despite building on previous concepts, has no raison d'etre per se. Of all possible paths building on previous section, why exactly this path? and why exactly in this form? This always happens in sequentially terse and polished treatments. Until now, it has not been a problem because I already had the motivation in mind, and it all linked together naturally. Even though my ultimate goal is fully understanding the three body problem, and hence integration and integrability, and hence calculus on manifolds, etc... There is a motivation hole between finishing chapter 2 in the Basics book and then next chapters in Analysis I leading the differentiation and anti-differentiation. Obviously, the path between the two is treated much more generally and completely than is needed in a 'shortest path' route, and that is fine, but we have to be aware of that, and should because of this not be hesitant to read forward, or search forward, for motivations of current study sections.

Note. Based on the note above, here are motivational results from the support book[11].

This motivates the need for passing to norms, which is obvious, but clearly, there is no harm in directly jumping to norms in general instead of Euclidean norm. Indeed, the fact that only some properties of the norm are needed is both interesting, and freeing. Since in a way, we are freed from the Pythagorean theorem's exclusivity and 'magic application' in Analysis, where the exclusivity would have been quite intriguing come to think of it.

> *The first section, IV.1, will introduce norms in n-dimensional spaces, which enable us to extend the definitions and theorems on convergence and continuity quite easily (Section IV.2). However, differential calculus (Sections IV.3 and IV.4) as well as integral calculus (Section IV.5) in several variables will lead to new difficulties (interchange of partial derivatives, of integrations, and of integrations with derivatives).*

Another quote about the drudgery of the current section, with an actual quote from Landau (it comes from [12]. Note the date)

> *This seems to resolve Theorem 1.8 in an elegant manner. But there remains much to do: we shall have to identify different rational Cauchy sequences that represent the same real number, define algebraic and order relations for these new objects, and finally we shall find the proof of Theorem 1.8 more complicated than we might have thought, because the terms sn in (1.14) may now*

themselves be real numbers, i.e., rational Cauchy sequences. All these details have been worked out in full detail by Landau (1930) in a famous book, where he admits himself that many parts are "eine langweilige Mühe".

(1.8) Theorem (Cauchy 1821). A sequence {sn} of real numbers is convergent (with a real number as limit) if and only if it is a Cauchy sequence.

This is a confirmation note about 'meta' (see note above) real numbers.

A new mathematical era began when Dedekind (about 1871) and Cantor (about 1875) considered sets of points as new mathematical objects.

And here is motivation for compact sets (and hence all the sections that lead to them).

(1.20) Remark. Compact sets are, by Definition 1.18, precisely the sets in which the Bolzano-Weierstrass theorem can be applied. Since this theorem is the basis for all deep results on uniform convergence, uniform continuity, maximum and minimum, Fréchet was not exaggerating (see quotation).

We have already pointed out and will recognize throughout this book the importance of compact sets. All those concerned with general analysis have seen that it is impossible to do without them. (Fréchet 1928, Espaces abstraits, p. 66)

Note the date of Frechet's quote

Note. It seems that I 'overdid' studying the dedekind cuts section (in a good way). After reading a lot (see notes above), and also redoing all the section proofs and pondering on them multiple times, I was ready. As a result, all the exercises became trivialities. There was one thing to be stressed again: the choice of steps to prepare for the field axiom proofs is very nice and elegant, providing short proofs of what otherwise would have been pages of cases with my initial (long time ago already) naive (but not wrong) attack.

Note. The definition and justification of infinites from Zakon quoted below now feels absolutely free from magic, I would even say intuitive and simple. This was made possible not only by all the notes above ('meta', 'black sheep', etc.), but also the study of Euclid's elements in comparison to Hilbert's axiomatic system and its philosophy (See my 'Notes on Euclid's Elements).

The Infinities. As we have seen, a setA $\neq \emptyset$ in E1 has a lub (glb) if A is bounded above (respectively, below), but not otherwise.
In order to avoid this inconvenient restriction, we now add to E1 two new objects of arbitrary nature, and call them "minus infinity" ($-\infty$) and "plus infinity" ($+\infty$) with the convention that $-\infty < +\infty$ and $-\infty < x < +\infty$ for all $x \in E1$.

Note. Zakon points out that ($+\infty$) is the maximum of E*, but we also notice a related and significant step that we reached in the last problem in the Dedekind cuts section:

> *Prove Dedekind's theorem: An ordered set is complete iff it has no gaps.*

Note how after *all* the work, we can finally in E1, substitute saying p=sup(A) for p=max(A) for a p determined by cut (A,B), p being any *real* number.

Note. More explanation of the relation between compactness and the basic notion of the limit of a sequence. The quote also illuminates the birth of generalizing the considered spaces to more than point in Rn, which is the reason why all these generalizations are explained in Zakon's book for the most general (and of course not historic) treatment possible (wikipedia). Curiously and sadly, there is no mention of Arzela-Ascoli in Hairer's book. The quote references Kline [13]. The quote also exposes flagrantly the high price of polish, terseness and non-historical treatment.

> ***In the 19th century, several disparate mathematical properties were understood that would later be seen as consequences of compactness. On the one hand, Bernard Bolzano (1817) had been aware that any bounded sequence of points (in the line or plane, for instance) has a subsequence that must eventually get arbitrarily close to some other point, called a limit point****. Bolzano's proof relied on the method of bisection: the sequence was placed into an interval that was then divided into two equal parts, and a part containing infinitely many terms of the sequence was selected. The process could then be repeated by dividing the resulting smaller interval into smaller and smaller parts until it closes down on the desired limit point. The full significance of Bolzano's theorem, and its method of proof, would not emerge until almost 50 years later when it was rediscovered by Karl Weierstrass.[1]*
>
> *In the 1880s, it became clear that results similar to the Bolzano–Weierstrass theorem could be formulated for spaces of functions rather than just numbers or geometrical points.* ***The idea of regarding functions as themselves points of a generalized space dates back to the investigations of Giulio Ascoli and Cesare Arzelà.[2] The culmination of their investigations, the Arzelà–Ascoli theorem, was a generalization of the Bolzano–Weierstrass theorem to families of continuous functions****, the precise conclusion of which was that it was possible to extract a uniformly convergent sequence of functions from a suitable family of functions. The uniform limit of this sequence then played precisely the same role as Bolzano's "limit point". Towards the beginning of the twentieth century, results similar to that of Arzelà and Ascoli began to accumulate in the area of integral equations, as investigated by David Hilbert and Erhard Schmidt. For a certain class of Green functions coming from solutions of integral equations, Schmidt had shown that a property analogous to the Arzelà–Ascoli theorem held in the sense of mean convergence—or convergence in what would later be dubbed a Hilbert space. This ultimately led to the notion of a compact operator as an offshoot of the general notion of a compact space. It was Maurice Fréchet who, in 1906, had distilled the essence of the Bolzano–Weierstrass property and coined the term compactness to refer to this general phenomenon.*
>
> *However, a different notion of compactness altogether had also slowly emerged at the end of the 19th century from the study of the continuum, which was seen as fundamental for the rigorous formulation of analysis. In 1870, Eduard Heine showed that a continuous function defined on a closed and bounded interval was in fact uniformly continuous. In the course of the proof, he made use of a lemma that from any countable cover of the interval by smaller open intervals, it was possible to select a finite number of these that also covered it. The significance of this lemma was recognized by Émile Borel (1895), and it was generalized to arbitrary collections of intervals by*

Pierre Cousin (1895) and Henri Lebesgue (1904). The Heine–Borel theorem, as the result is now known, is another special property possessed by closed and bounded sets of real numbers.

This property was significant because it allowed for the passage from local information about a set (such as the continuity of a function) to global information about the set (such as the uniform continuity of a function). This sentiment was expressed by Lebesgue (1904), who also exploited it in the development of the integral now bearing his name. Ultimately the Russian school of point-set topology, under the direction of Pavel Alexandrov and Pavel Urysohn, formulated Heine–Borel compactness in a way that could be applied to the modern notion of a topological space. Alexandrov & Urysohn (1929) showed that the earlier version of compactness due to Fréchet, now called (relative) sequential compactness, under appropriate conditions followed from the version of compactness that was formulated in terms of the existence of finite subcovers. ***It was this notion of compactness that became the dominant one, because it was not only a stronger property, but it could be formulated in a more general setting with a minimum of additional technical machinery, as it relied only on the structure of the open sets in a space.***

The more detailed related quote from Kline. Here we also notice the totally unexpected origins of Linear Algebra. The subtle hints of Hefferon's book now come into the light.

The absttract theory of functionals was initiated by Volterra in work concerned with the calculus of variations. His work on functions of lines (curves), as he called the subject, covers a numbers of papers.

A function of lines was, for Volterra, a real-valued function F whose values depend on all the values of functions y(x) defined on some interval [a,b]. The functions themselves were regarded as points of space for which the neighborhood of a point and the limit of a sequence of points can be defined. For the functionals F[y(x)] Volterra offered definitions of continuity, derivative and differential.

However, these definitions were not adequate for the abstract theory of the calculus of variations and were superseded. His definitions were in fact criticized by Hadamard.

Even before Volterra commenced his work, the notion that a collection of functions y(x) all defined on some interval be regarded as points of a space had already been suggested. Riemann, in his thesis, spoke of a collection of functions forming a connected closed domain (of points of a space).

Giulio Ascoli (1843-1896) and Cesare Arzela sought to extend to sets of functions Cantor's theory of sets of points, and so regarded functions as points of a space.

Arzela also spoke of functions of lines. Hadamard suggested at the First International Congress of Mathematicians in 1897 that curves can be considered as points of a set.

He was thinking of the family of continuous functions defined over [0,1], a family that arose in his work on partial differential equations. Emile Borel made the same suggestion for a different purpose, namely, the study of arbitrary functions by means of series.

Hadamard, too, undertook the study of functionals on behalf of the calculus of variations. The term functional is due to him. According to Hadamard, a functional U[y(t)] is linear when $y(t)=\lambda_1 y_1(t)+\lambda_2\ y_2(t)$*, wherein a1 and a2 are constants, then* $U[y(t)] = \lambda_1 U[y_1(t)] + \lambda_2\ U[y_2(t)]$*.*

The first major effort to build up an abstract theory of function spaces and functionals was made by Maurice Frechet (1878-1973), a leading French professor of mathematics, in his doctoral thesis of 1906. In what he called the functional calculus he sought to unify in abstract terms the ideas in the work of Cantor, Volterra, Arzela, Hadamard and others.

Note. In the section on upper and lower limits, I am constantly imagining what seems to be described in later sections as accumulation points, and this image comes very naturally. Since as soon as we appreciate upper and lower limits, it is easy to imagine a sequence that infinitely repeats three distinct reals. We feel that there are three 'trivial' accumulation points that we could single out. On a rougher level, any number of accumulation points that is not exactly one, will forcefully result in distinct upper and lower limits, the upper limit being the largest accumulation point, and the lower limit being the lowest. On an even rougher level, the limit defined in terms of a neighborhood only ever exists if there is only one accumulation point, and is that point.

Note. It is intuitive to understand that there would be a relation between the two definitions of the limit in the section (explained exactly with the section's theorem). Since, on a rough level, the neighborhood definition is *bidirectional,* whereas the upper and lower limits, being a sup and and inf are *unidirectional.*

Note. Although formally, the upper and lower limits are defined as an inf of sups and sup of infs, it is convenient to think of them as a sup and an inf *at infinity*. This also helps remember the useful corollaries of the section that undoubtedly allow for shorter proofs. If infinitely many x_n are larger than some b, then b *carries* the supremum at infinity (pulls for the inf version). If all (but finitely many) x_n are are less than some a, then a *caps* the supremum at infinity. The converses also holds and are naturally proven by contradiction. Attention must be made to the exact kinds of inequality used. For the direction explained above, the inequalities are not strict, because b could *carry* the supremum, but still be equal to the upper limit as is not unusual for infs (and sups); the iconic example of zero as the inf of positive numbers is always a good example. The converses then use strict inequalities.

Note. A good way to reason fast about the distinction between *finitely many*, *infinitely many*, and *all but finitely many*, is to be aware of the *split* they produce. Each splits the elements in the sequence (or set in general) into two groups. One in which the predicate P in question is true and one in which it is false.

- *Finitely many* are true, infinitely many are false.
- *Infinitely many* are true, the rest (finite or infinite) is false.
- *All* are true, but for *finitely many* which are false. In other words, *Finitely many* are false, infinitely many are true.

Clearly, the first and last are of a stronger nature, as they make a definite statement about the finiteness of the *false* group.

Note. Despite trying to reduce all the above concepts to ideas that are easy and fast to reason about, finessing the reasoning methods can only happen with a lot of practice and is all about being able to precisely state finitely certain truths about infinite sets or processes, especially about inclusion, exclusion and existence.

Note. If anything, the problems in the section are a lesson in corollary development. They are all provable without them, but the result would be too long. How would one detect a situation where corollaries are needed? One idea is to think about creating them when the fact to be proven seems plausible (or is given in a problem to prove and hence true) but the reasoning path seems longer than can be mentally handled. In this case, a mental proof can not suffice, so partial results have to be gathered on paper as corollaries, and the problem should be reconsidered in the light of corollaries. This is however still not a method as the forward-thinking *third-eye* has to guide the choice of useful partial results.

Note. Spivak provides a very subtle but 'better' statement of the intersection of *nested intervals,* in a way that fits more closely the idea of finite statements about infinite processes. Instead of writing $\cap^{\infty}_{n=1} [a_n, b_n]$, he says: *Prove that there is a point x which is in every interval* [9 (Pg. 120)]. Another good idea is that he immediately asks: *Show that this conclusion is false if we consider open intervals instead of closed intervals.* Which boils down to the fact that open intervals in E^1 can never contain only one point, which would have to be the case for the sought after intersection.

Note. Here is a nice quote about how Physics (the vibrating string) showed that the simpler algebraic functions could no longer be used as a model for functions in general, in terms of continuity, expansion, existence of derivatives and integrals.

> *In view of the almost complete absence of any foundations, how could the mathematicians proceed with the manipulation of the variety of functions? In addition to their great reliance upon physical and intuitive meanings, they did have a model in mind- the simpler algebraic functions, such as polynomials and rational functions. They carried over to all functions the properties they found in these explicit, concrete functions: continuity, the existence of isolated infinities and discontinuities, expansion in power series, and the existence of derivatives and integrals. But when they were obliged, largely through the work on the vibrating string, to broaden the concept of function, as Euler put it, to any freely drawn curves (Euler's mixed of irregular or discontinuous functions), they could no longer use the simple functions as a guide. And when the logarithmic function had to be extended to negative and complex numbers, they really proceeded without any reliable basis at all; this is why arguments on such matters were common. The rigorization of the calculus was not achieved until the nineteenth century.*[13]

I am not sure yet why the vibrating string studies necessitated such functions. Maybe because the string could be in a random (freely drawn) initial position? I will have to find out.

Note. I thought of justifying the need for a historical study using *causal prerequisites (the causes of why the solution such and such, the problems that lead to forming it)*, in contrast to 'mathematical' prerequisites. Analysis as presented by Zakon --the real number system, limits, compactness, etc.-- carries the solutions to all the historic problems so well documented in Kline[13]. The mathematical prerequisites of his presentation are none. Are the causal prerequisites unimportant? Ultimately, one can say that the issues the mathematicians were struggling with and could not solve are intrinsically different from the problems one should and will be dealing with if one wants to do research in this day and age. However these is still value in studying the causal prerequisites, and not only for learning from the greats, which could be judged at not being a good enough reason. There is so much subtlety hidden in the proposed solution, that one will philosophically stumble when encountering some of the subtle problems, even if technically he is supposed to have learned how to deal with it technically. At least an allusion to the problems of the past is a must, and this should be an integral part of the explanation, and not some unspectacular exercise. Understanding mathematics might be arguably separated from understanding human thought, but practically they go hand in hand. Understanding the reflexes of human thought confronted to those historical problems is understanding one's own preliminary thoughts, understanding the contrast between these thoughts that every human will initially have, and each philosophical and technical stepping stone, that allowed us to refine our mathematical (and ultimately also common) language in order to analyse the topics at hand. If this is not the essence of 'understanding' mathematics, then what is.

Note. My best option to handle both the need to advance in the way of standard mathematics towards my goals (e.g modern variational calculus as applied to physics) while at the same time rooting myself in the history of mathematical thought (made even more important after the immense benefits of studied Euclid, reading Heath and Kline), I must have the discipline to make the mental thread between the two things thinner. To be able to proceed in each (as a student in the former, since by Greek definition, mathematics must be learned; as a thinker in the latter) without fully relying on progress in the other. I have noticed that Kline[13] will really be of phenomenal help. And again and again Hairer as well. It is encouraging that the path from Euclid to 17th century calculus is not that insurmountable. The greek Conic sections being the next major step in the way. This tell me that the current split plan of study is not a bad one.

Appendix

History of Topology

Quotes from the book 'History of Topology' by I.M James.

> *The twin problems of defining dimension and proving its invariance were the primary influences on the creation and growth of modern topological dimension theory. However, the cosmological problem of explaining the dimension number of physical space also was a secondary motive.*
> *The history of modern dimension theory reveals a wide variety of definitions and theories. No single definition and theory can be regarded as uniquely correct and, contrary to the beliefs of the earliest workers, we cannot expect a single definition of dimension to reveal the "true essence" of the concept. The search for the holy grail, the single, universally acceptable definition has proved illusory. No single theory of dimension is exclusively at the centre of the mathematical stage for all time. In modern times three definitions of topological dimension are regarded as important, not counting the metrical dimension (Hausdorff/Besicovitch) which has been significant over the last twenty years in its connection with fractals. These are the "small inductive dimension" (Menger-Urysohn), the "large inductive dimension" (Brouwer-Cech) and the "covering dimension" (Cech-Lebesgue) [21]. Brouwer was perhaps the pivotal figure in the development of topology in the twentieth century. After 1913, Brouwer's contributions to topology were few but he still remained an authoritative presence. During the period 1925-1926, Alexandroff, Menger, and the Austrian Leopold Vietoris (b. 1891) visited Brouwer in Amsterdam (Vietoris is now the grand old man of Austrian mathematics). Through Menger, Witold Hurewicz became a mathematical assistant to Brouwer, and Hans Freudenthal, attracted by Brouwer's philosophical work, became another. Brouwer also influenced the American mathematician J.W. Alexander. In 1922 Alexander generalised the Jordan curve to higher dimensions, a result now known as Alexander duahty. Brouwer inspired Erhard Schmidt (1876-1959), and through Schmidt's lectures, Heinz Hopf (1894-1971) was drawn into the subject. With such strong personalities as Brouwer, Menger, and Alexandroff involved, the sweet reasonableness of the earlier Cantor-Dedekind dialogue proved a rarity in the history of dimension theory.*

> *A history of dimension theory after the 1920s, after it became increasingly abstract and axiomatic, is outside the scope of this survey. Dimension theory as a collection of mathematical theories has grown rapidly during the twentieth century - a brief outline of technical developments in the 1920s and 1930s is given in [21]. In many ways its growth illustrates the kinds of development which mathematical theories often follow. Principally it advanced through the efforts of mathematicians to solve significant problems. New directions opened up, and the theories became more abstract. For instance, in 1932 Alexandroff extended the theory towards homology theory and algebraic topology in his paper "Dimensionstheorie". Other mathematicians contributed to dimension theory, most notably L.A. Tumarkin (b. 1904), Lev Pontryagin (1908-1988), Georg Nobeling (b. 1907) and Eduard Cech (1893-1960). In 1941 Witold Hurewicz and Henry Wallman (1915-1992), in dealing very succinctly with separable metric topological spaces set a standard for exposi-*

tion in their classic Dimension Theory [14].
At the beginning of this article it was noted that space was generally accepted as having three dimensions and this notion was relatively unproblematic. The opening lines of W. Hurewicz and H. Wallman's Dimension Theory suggests an appropriate note on which to end [14, p. 3]:

Of all the theorems of analysis situs, the most important is that which we express by saying that space has three dimensions. It is this proposition that we are about to consider, and we shall put the question in these terms: when we say that space has three dimensions, what do we mean?

In the early 19th century we find diverse steps towards a generalization of geometric language to higher dimensions. But they were still of a tentative and often merely metaphorical character. The analytical description of dynamical systems in classical mechanics was a field in which, from hindsight, one would expect a drive towards and a growing awareness of the usefulness of higher dimensional geometrical language.^ But the sources do not, with some minor exceptions, imply such expectations. Although already Lagrange had used the possibility to consider time as a kind of fourth dimension in addition to the three spatial coordinates of a point in his Mecanique analitique (1788) and applied a contact argument to function systems in 5 variables by transfer from the 3-dimensional geometrical case in his Theorie des fonctions analytiques (1797, Section 3.5.25), these early indications were not immediately followed by others.
Not before the 1830-s and 1840-s do we find broader attempts to generalize geometrical language and geometrical ideas to higher dimensions: Jacobi (1834), e.g., calculated the volume of n-dimensional spheres and used orthogonal substitutions to diagonalize quadratic forms in n variables, but preferred to avoid explicit geometrical language in his investigations. Cayley's Chapters in the analytical geometry ofn dimensions (1843) did use such explicit geometrical language - but still only in the title, not in the text of the article.
*It was the following decade about the middle of the century which brought the change. In a short time interval we find a group of authors using and exploring conceptual generalizations of geometrical thought to higher dimensions, without in general knowing about each other. Among them was Grassmann with his Lineale Ausdehnungslehre (1844) containing an expUcit program for a new conceptual foundation for geometry on n-dimensional (lin-ear) extensional quantities,^ Pliicker with his System der Geometrie des Raumes (1846) and 4-dimensional Hne geometry in classical 3-space, and, in a certain respect most elaborated among these attempts, Schlafli with his Theorie der vielfachen Kontinuitdt (1851/1901), which was published only posthumously.^*
Also leading mathematicians like Cauchy and Gauss started to use geometrizing language in M" in publications (Cauchy, 1847) or lecture courses (GauB, 1851/1917). Gauss, in his lecture courses, even used the vocabulary of {n — kydimensional manifolds (Mannigfaltigkeiten), but still restricted in his context to affine subspaces of the ^-dimensional real space (GauB, 1851/1917, pp. 477ff.). There is no reason to doubt that Riemann got at least some vague suggestion of how to generalize the basic conceptual frame for geometry along these lines from Gauss and developed it in a highly independent way.

When Riemann presented his ideas on a geometry in manifolds the first time to a scientific audience in his famous Habilitationsvortrag (Riemann, 1854), he was completely aware that he was working in a border region between mathematics, physics, and philosophy, not only in the sense of the pragmatic reason that his audience was mixed, but by the very nature of his exposition.^ There was no linguistic or symboUcal frame inside mathematics, which he could refer to, even only to formulate a general concept of manifold. So he openly drew on the resources of contemporary idealist, dialectical philosophy, in his case oriented at J.F. Herbart, to generalize the classical concept of extended magnitude/quantity for geometry and to "construct" the latter as only one specification from a more general concept.^ Basic to such a construction was, so Riemann explained to his audience, the presupposition of any "general concept" which allows in a logical sense precise individual determinations. From the extensional point of view such a concept would form a manifold and the individual modes of determination were to be considered, as Riemann explicitly stated, as the elements or the points of the manifold with either "discrete" or "continuous" transition from one to the other. Thus Riemann sketched the draft for a conceptual starting point for what later was to become general set theory (discrete manifolds)^ and topology (continuous manifolds).

*Such concepts would gain mathematical value only if a sufficiently rich structure of (real or complex valued) functions on the manifold is available. Then it should be possible to describe the specification of points by the values of n properly chosen functions in a locally unique way (local coordinate system). That a change of coordinates would lead to locally invertible differentiable real functions, was not made explicit by him, but was to be understood from the context by careful listeners or readers. The distinction between local simplicity of manifolds, because of the presupposition of local coordinate systems, and globally involved behaviour was indicated by Riemann, but not particularly emphasized during the talk, although in other pubhcations and manuscripts it was.^*

The reception and assimilation of Riemann's concept of manifold to the mathematics of the 19th century was slow and inhibited by severe conceptual problems. Of course it was difficult to understand what a manifold in general should be. The easiest way was to translate it as a "number manifold" in the 1870-s and later. At that time the former real quantities had been arithmetically reconstructed by Meray, Cantor, Dedekind, and Weierstrass, and it appeared as perfectly clear to talk about concretely given submanifolds of R'^ or of projective spaces Pfn^ or P^C. Such submanifolds were in the easiest approach defined by inequahties as m-dimensional (usually connected) subsets in the works of Beltrami (1868a, 1868b) Helmholtz (1868), and even of the young Klein during his investigations on non-Euclidean geometry and the Erlangen program (1871).

That was of course a reduction of Riemann's intention and suppressed the distinction between local simplicity and global complexity of manifolds. That global behaviour was an essential ingredient for Riemann's concept, was most clearly understood in the 1860-s and 1870-s in the special context of geometric function theory and the dissemination of knowledge about the topology of Riemann surfaces (Liiroth, Clebsch, Neumann, Clifford et al.) An additional aspect was the problem of compactification of geometrical objects "in the infinite", which in a discussion between Schlafli and Klein was realized, when they debated the difference between one-point compactification of the plane Uke C to PiC and Hne compactification of E^ to P2R and its topological consequences (Schlafli, 1872; Klein, 1873, 1874-1876).

Early axiomatic attempts for two-dimensional manifolds

Topological spaces on different levels of generalization were analyzed in different approaches and with varying degrees of precision in the rise of modern mathematics in the early 20th century. During the last three decades of the 19th century Cantor had developed his theory of point sets in R" in the framework of general set theory. He himself was shocked to realize that bijective maps between real continua of different dimensions can be conceived, and even Dedekind's comforting conviction that more specific maps, in this case bijective and (bi-)continuous ones, would respect the invariance of dimension left the problem to prove (or disprove) such a conjectured invariance. Naive assumptions from space intuition were particularly deceptive in this field; that became even clearer about 1890 when Peano published his example of a "spacefilling" curve with the surprising effect, that the lack of injectivity would even for continuous maps not necessarily lead to a decrease of dimension (or keep it at most invariant), but could as well increase it. Early attempts by Liiroth, Thomae, Netto, and Cantor himself, to prove the invariance of dimension under bijective continuous maps, turned out to contain unclosable gaps and again (as in the case of the continuity proof for uniformization) it was only Brouwer who surmounted the difficulties and indeed proved the correctness of Dedekind's suggestion (Brouwer, 191 la).^^ About the turn of the century two methodological strategies for clarifying the concept of manifold were formed and sketched, an axiomatic one proposed by Hilbert, taken up by Weyl (about 1913), Hausdorff, H. Kneser, and VeblenAVhitehead, and a constructive one proposed by Poincare, taken up by Dehn/Heegard, Tietze, Steinitz, Brouwer, Weyl (after 1920), Vietoris, van Kampen and others.

^^ I deliberately use Brouwer's original terminology and do not write R", as Brouwer's terminology leaves the interpretation of the number continuum open. It can be interpreted by classical real numbers, Brouwer's inituitionistic real continuum, or even (later) by Weylian reals W of 1918.

“Finally the "modern " axiomatic concept

There was, of course, still another line of research, more closely linked to differential geometry, where manifolds played an essential role, and purely topological aspects (independently of whether continuous, combinatorial, or homological ones) did not suffice and still needed elaboration. In North America Oswald Veblen and his students formed an active center in both fields of topology and modern geometry. Veblen and his student J.H.C. Whitehead, coming from (and going back to) Oxford, brought the axiomatization of the manifold concept to a stage which stood up to the standards of modern mathematics in the sense of the 20th century (Veblen and Whitehead, 1931, 1932). Veblen was an admirer of the Gottingen tradition of mathematics, in particular, F. Klein and D. Hilbert, and cooperated closely with H. Weyl, the broadest representative of his own generation from the Klein and Hilbert tradition. Veblen and J.H.C. Whitehead combined a view of the central importance of structure groups for geometry (generaUzing the Erlanger program) with Hubert's embryonic characterization of manifolds by coordinate systems; and they took care that the topologization of the underlying set would satisfy Hausdorff's axioms for a topological space.

...
Whitehead and Veblen presented their axiomatic characterization of manifolds of class G first in a research article in the Annals of Mathematics (Veblen and Whitehead, 1931) and in the final form in their tract on the Foundations of Differential Geometry (Veblen and Whitehead, 1932). Their book contributed effectively to a conceptual standardization of modern differential geometry, including not only the basic concepts of continuous and differentiable manifolds of different classes, but also the "modern" reconstruction of the differentials dx = {Ax\,..., dxn) as objects in tangent spaces to M}^^ Basic concepts like Riemannian metric, affine connection, holonomy group, covering manifolds, etc. followed in a formal and symboUc precision that even from the strict logical standards of the 1930-s there remained no doubt about the wellfoundedness of differential geometry in manifolds. Moreover they made the whole subject conceptually accessible to anybody acquainted with the language and symbolic practices of modern mathematics." [3]

Some Definition Attempts for the Real Numbers

All quotes below originate from [4] except if mentioned otherwise.
Note. An equally nice chronology can be found in the 'Princeton Companion to Mathematics' starting page 126.

Cauchy, not a definition, but a significant precursor:

... though Cauchy implicitly assumed several forms of the completeness axiom for the real numbers, he did not fully understand the nature of completeness or the related topological properties of sets of real numbers or of points in space. ... Cauchy did not have explicit formulations for the completeness of the real numbers. Among the forms of the completeness property he implicitly assumed are that a bounded monotone sequence converges to a limit and that the Cauchy criterion is a sufficient condition for the convergence of a series. Though Cauchy understood that a real number could be obtained as the limit of rationals, he did not develop his insight into a definition of real numbers or a detailed description of the properties of real numbers.

Hamilton:

One of the first people to attempt to give a rigorous definition of the real numbers was Hamilton. Perhaps, if one thinks about it, it is logical that he would be interested in this since his introduction of the quaternions had shown that there were new previously unstudied number systems. In fact came close to the idea of a Dedekind cut, as Euclid had done in the Elements, but failed to make the idea into a definition (again Euclid had spotted the property but never thought to use it as a definition). For a number a he noted that there are rationals a', a", b', b", c', c", d', d", ... with

$a' < a < a''$ $b' < a < b''$ $c' < a < c''$ $d' < a < d''$ ...

but he never thought to define a number by the sets {a', b', c', d', ... } and {a", b", c", d", ... }. He tried another approach of defining numbers given by some law, say x->x2. Hamilton writes:-

If x undergoes a continuous and constant increase from zero, then will pass successively through every state of positive ration b, and therefore that every determined positive

> *ration b has one determined square root √b which will be commensurable or incommensurable according as b can or cannot be expressed as the square of a fraction. When b cannot be so expressed, it is still possible to approximate in fractions to the incommensurable square root √b by choosing successively larger and larger positive denominators ...*

One can see what Hamilton is getting at, but much here is without justification - can a quantity undergo a continuous and constant increase. Even if one got round this problem he is only defining numbers given by a law. It is unclear whether he thought that all real numbers would arise in this way.

Unpublished Dedekind and Weierstrass:

When progress came in giving a rigorous definition of a real number, there was a sudden flood of contributions. Dedekind worked out his theory of Dedekind cuts in 1858 but it remained unpublished until 1872. Weierstrass gave his own theory of real numbers in his Berlin lectures beginning in 1865 but this work was not published. The first published contribution regarding this new approach came in 1867 from Hankel who was a student of Weierstrass.

Hankel, not a definition, but a significant point of view:

Hankel, for the first time, suggests a total change in out point of view regarding the concept of a real number:-

> *Today number is no longer an object, a substance which exists outside the thinking subject and the objects giving rise to this substance, an independent principle, as it was for instance for the Pythagoreans. Therefore, the question of the existence of numbers can only refer to the thinking subject or to those objects of thought whose relations are represented by numbers. Strictly speaking, only that which is logically impossible* (i.e. *which contradicts itself) counts as impossible for the mathematician.*

In his 1867 monograph Hankel addressed the question of whether there were other "number systems" which had essentially the same rules as the real numbers

Frege later

Meray, note that this is the one that first comes to mind once one learns the completeness axiom following Zakon.

Two years after the publication of Hankel's monograph, Méray published Remarques sur la nature des quantités in which he considered Cauchy sequences of rational numbers which, if they did not converge to a rational limit, had what he called a "fictitious limit". He then considered the real numbers to consist of the rational numbers and his fictitious limits.

Heine,

Note. I did not reach that in Zakon's book yet, because he starts with the completeness axiom and then follows with Dedekind Cuts (I really like his point of view on a polished path and planting the completeness axiom in the reader's head as the first contact. since it really feels like the most 'atomic'), but I understand it because I spent time pondering about the cuts and why not something along Meray was used and concluded that it is because the subtleties of the definition of the cuts, requiring all bounds on both sides nicely guarantee the 'uniqueness', each cut is ‘unique’, whereas this is not the case for ideas like Meray’s, or my first idea after the completeness axiom, where multiple bounded sets could have the same sup, also the issue with some having a sup and some having an inf. is elegantly bypassed, or even Heine’s where the ‘uniqueness’ is solved using equivalence classes.

> *Three years later Heine published a similar notion in his book Elemente der Functionenlehre although it was done independently of Méray. It was similar in nature with the ideas which Weierstrass had discussed in his lectures. Heine's system has become one of the two standard ways of defining the real numbers today. Essentially Heine looks at Cauchy sequences of rational numbers. He defines an equivalence relation on such sequences by defining*
> *a1 ,a2 ,a3 ,a4 ,...and b1,b2 ,b3 ,b4 ,...*
> *to be equivalent if the sequence of rational numbers a1 - b1, a2 - b2 , a3 - b3 , a4 - b4 , ... converges to 0. Heine then introduced arithmetic operations on his sequences and an order relation. Particular care is needed to handle division since sequences with a non-zero limit might still have terms equal to 0.*

Cantor:

> *Cantor also published his version of the real numbers in 1872 which followed a similar method to that of Heine. His numbers were Cauchy sequences of rational numbers and he used the term "determinate limit".*

One escape to reconcile the slow departure of real numbers from (geometric, metric) quantity was provided by Cantor, by an axiom!

> *It was clear to Hankel (see the quote above) that the new ideas of number had suddenly totally changed a concept which had been motivated by measurement and quantity. Similarly Cantor realised that if he wants the line to represent the real numbers then he has to introduce an axiom to recover the connection between the way the real numbers are now being defined and the old concept of measurement. He writes about a distance of a point from the origin on the line:-*
>
> > *If this distance has a rational relation to the unit of measure, then it is expressed by a rational quantity in the domain of rational numbers; otherwise, if the point is one known through a construction, it is always possible to give a sequence of rationals a1 , a2 , a3 ,..., an , ... which has the properties indicated and relates to the distance in question in such a way that the points on the straight line to which the distances a1 , a2 , a3 , ..., an , ... are assigned approach in infinity the point to be determined with increasing n. ... In order to complete the connection presented in this section of the domains of the quantities defined [his determinate limits] with the geometry of the straight line, one must add an axiom which simple says that every numerical quantity also has a determined point on the straight line whose coordinate is equal to that quantity, indeed, equal in the sense in which this is explained in this section.*

And here is Dedekind’s which as we mentioned, comes first in the book:

> *... Dedekind had worked out his idea of Dedekind cuts in 1858. When he realised that others like Heine and Cantor were about to publish their versions of a rigorous definition of the real numbers he decided that he too should publish his ideas. This resulted in yet another 1872 publication giving a definition of the real numbers.*

Dedekind considered all decompositions of the rational numbers into two sets A1 , A2 so that a1 < a2 for all a1 in A1 and a2 in A2. He called (A1, A2) a cut. If the rational a is either the maximum element of A1 or the minimum element of A2 then Dedekind said the cut was produced by a. However not all cuts were produced by a rational. He wrote:-

> *In every case in which a cut (A1, A2) is given that is not produced by a rational number, we create a new number, an irrational number a, which we consider to be completely defined by this cut; we will say that the number a corresponds to this cut or that it produces the cut*

He defined the usual arithmetic operations and ordering and showed that the usual laws apply.

More justifications to the ‘departure’ came from Thomae, a colleague of Cantor and Heine:

He claimed that the real numbers defined in this way had a right to exist because:-

> *... the rules of combination abstracted from calculations with integers may be applied to them without contradiction.*

Frege was not convinced:

Frege, however, attacked these ideas of Thomae. He wanted to develop a theory of real numbers based on a purely logical base and attacked the philosophy behind the constructions which had been published. Thomae added further explanation to his idea of "formal arithmetic" in the second edition of his text which appeared in 1898:-

> *The formal conception of numbers requires of itself more modest limitations than does the logical conception. It does not ask, what are and what shall the numbers be, but it asks, what does one require of numbers in arithmetic.*

Frege was still unhappy with the constructions of Weierstrass, Heine, Cantor, Thomae and Dedekind. How did one know. he asked, that the constructions led to systems which would not produced contradictions? He wrote in 1903:-

> *This task has never been approached seriously, let alone been solved.*

Frege, however, never completed his own version of a logical framework. His hopes were shattered when he learnt of Russell's paradox.

Hilbert’s definition is closest to the book. All the concepts leading to the cuts in the book are given by his axiomatic system, except for the formulation of the completeness axiom, which Zakon probably had to simplify for us by taking a simpler route (a guess) that is not based on mathematical logic since logic it is not a prerequisite of the book:

Hilbert had taken a totally different approach to defining the real numbers in 1900. He defined the real numbers to be a system with eighteen axioms. Sixteen of these axioms define what today we call an ordered field, while the other two were the Archimedean axiom and the completeness axiom. The Archimedean axiom stated that given positive numbers a and b then it is possible to add a to itself a finite number of times so that the sum exceed b. The completeness property says *that one cannot extend the system and maintain the validity of all the other axioms. This was totally new since all other methods built the real numbers from the known rational numbers. Hilbert's numbers were unconnected with any known system. It was impossible to say whether a given mathematical object was a real number. Most seriously, there was no proof that any such system actually existed. If it did it was still subject to the same questions concerning its consistency as Frege had pointed out.*

The Current State of the Common Paradoxes

All quotes below originate from [4] except if mentioned otherwise.

A summary of the problem:

In his list of problems which he proposed to the International Congress of Mathematicians at Paris in August 1900, Hilbert stated that one of the most pressing issues for the foundations of mathematics was a proof of the consistency of arithmetic. He attempted to solve this himself but was unsuccessful. It was one thing trying to prove arithmetic consistent, but it was known that set theory led to paradoxes. These paradoxes worried many mathematicians and they felt that the foundations of mathematics needed to be built on a logical foundation which did not contain inherent contradictions. Three major paradoxes were due to Burali-Forti in 1897, Russell in 1902, and Richard in 1905. The first of these was derived from the fact that the ordinal numbers themselves formed an ordered set whose order type had to be an ordinal number. Russell's paradox is the well-known one relating to the set of all sets which do not contain themselves as an element, and Richard's paradox we have explained above. Solutions proposed by some mathematicians would only allow mathematics to treat objects which could be constructed. Poincaré (1908) and Weyl (1918) complained that analysis had to be based on a concept of the real numbers which eliminated the non-constructive features. Weyl argued in this 1918 work that analysis should be built on a countable continuum. It was the uncountable, and so non-constructible, aspects of the real line which Weyl felt caused problems.

A summary of the general state of affairs:

Gödel proved some striking theorem in 1930. He showed that a formal theory which includes the arithmetic of the natural numbers had to lead to statements which could neither be proved nor disproved within the theory. In particular the consistency of arithmetic was unprovable unless one used a higher order system in which to create the proof, the consistency of this system being equally unprovable. In 1936 Gentzen proved arithmetic consistent, but only by using transfinite methods which were less accepted than arithmetic itself. Although this topic of research is still an active one, most mathematicians accept the uncountable world of Cantor and the non-constructive system of real numbers.

Richard's Paradox is described as follows:

We'll construct a certain real number which, although historically not one that was looked at, will let us understand some of the questions that arose. Let us start with the 100 two digit numbers. A simple code will let us translate these into letters, 00 become a, 01 become b, ... , 25 becomes z, *26 becomes A, 27 becomes B, ... , 51 becomes Z, then code all the punctuation marks, and then make all the remaining numbers up to 99 translate to an empty space. Now create a number, say c, starting from the 100 2-blocks.*

c = 0.01020304050607080910111213141516171819202122232425...
Then continue with the 10000 pairs of 2-blocks 0000, 0001, 0002, ..., 0099, 0100, 0101, ... Then the 1000000 triples of 2-blocks etc. We can represent c as a point on a line segment of length 1. Yet every English sentence ever written or ever to be written, occurs in the decoding of c into letters. For example "one third" has 9 characters so will be decoded from c around 1018 digits after the decimal point. This article is there, both with the misprints which inevitably occur and a corrected version is there (but one has to go rather a long way to the right of the decimal point to find it!). The whole of Shakespeare is there, as is every book yet to be written, etc!

Let us use c to describe a paradox which was discovered in 1905. The first thing to notice is that all descriptions of real numbers in English (let us forget about words in other languages) must appear in c, since every possible sentence occurs in c. For example, "one third" will occur as we have noted, as will "the base of natural logs", and "the ratio of the circumference of a circle to its diameter", etc. This will enable us to explain Richard's paradox, discovered by Jules Richard in 1905. There are only a countable number of such descriptions of real numbers in English so all real numbers (except a tiny countable subset) can never be described in English. However, this is not the paradox. Richard obtained that by using Cantor's diagonal argument. Create a real number r, say, as follows:

If the n-th block of c translates into a description of a real number r(n) then set the n-th digit of r to be different to the n-th digit of r(n). If the n-th block of c does not describe a real number (most of course will not even be meaningful in English) then set the n-th digit of r(n) to be 1.

Now the real number r cannot be described in English, since it differs by construction from every real number which can be described in English. That is a bit worrying. Even worse, of course, is the fact that we have just described r in English in the previous paragraph! If Richard's paradox tells us anything then perhaps it is a warning not to use English (or any other language for that matter) when we are doing mathematics.

The current state seems to be:

The proposed definition of the new real number r clearly contains a finite string of characters, and hence it appears at first to be a definition of a real number. However, the definition refers to definability-in-English itself. If it were possible to determine which English expressions actually do define a real number, and which do not, then the paradox would go through. Thus the resolution of Richard's paradox is that there is no way to unambiguously determine exactly which English sentences are definitions of real numbers (see Good 1966). That is, there is no way to describe in a finite number of words how to tell whether an arbitrary English expression is a definition of a real number. This is not surprising, as the ability to make this determination would also imply the ability to solve the halting problem and perform any other non-algorithmic calculation that can be described in English. [6]

and

Another viewpoint on Richard's paradox relates to mathematical ***predicativism****. In this view, the real numbers are defined in stages, with each stage only making reference to previous stages*

and other things that have already been defined. From a predicative viewpoint it is not valid to quantify over all real numbers in the process of generating a new real number, because this is believed to lead to a vicious-circle problem in the definitions. Set theories such as ZFC are not based on this sort of predicative framework, and allow impredicative definitions.
Richard (1905) presented a solution to the paradox from the viewpoint of predicativisim. Richard claimed that the flaw in the paradoxical construction was that the expression for the construction of the real number r does not actually unambiguously define a real number, because the statement refers to the construction of an infinite set of real numbers, of which r itself is a part. Thus, Richard says, the real number r will not be included as any r_n, because the definition of r does not meet the criteria for being included in the sequence of definitions used to construct the sequence r_n. Contemporary mathematicians agree that the definition of r is invalid, but for a different reason. They believe the definition of r is invalid because there is no well-defined notion of when an English phrase defines a real number, and so there is no unambiguous way to construct the sequence r_n.
*Although Richard's solution to the paradox did not gain favor with mathematicians, predicativism is an important part of the study of the foundations of mathematics. Predicativism was first studied in detail by Henri **Poincaré** and Hermann Weyl in Das Kontinuum, where they showed that much of elementary real analysis can be conducted in a predicative manner starting with only the natural numbers. More recently, predicativism has been studied by Solomon **Feferman**, who has used proof theory to explore the relationship between predicative and impredicative systems.* [6]

Note that the items in bold are the path I was embracing and reading about in my previous studies of the paradoxes, Reading Bishop famous work was the end-point. At the time I decided to postpone this until I fully understand standard real analysis because otherwise, the subtleties of Bishop's book would be totally lost on me. Another obvious reason would be that this is a luxury topic at the moment.

Finally, the quote also contains an important sentence important that settles a question I had:

Set theories such as ZFC are not based on this sort of predicative framework, and allow impredicative definitions. [6]

Borel did not find any paradoxes, but we list him nevertheless, his *normal numbers* are interesting (and very intuitive) as a hint that the real numbers are in some way, nevertheless well behaved.

Emile Borel introduced the concept of a normal real number in 1909. His idea was to provide a test as to whether the digits in a real number occurred in the sort of way they would if we chose each one at random. First assume that we have a real number written in base 10, that is a decimal expansion. Then if it is a "random" number the digit 1 should occur about 1/10 of the time

so, if we denote by N(1,n) the number of times 1 occurs in the first n decimal digits, then N(1,n)/n should tend to 1/10 as n tends to infinity. Similarly for the all digits i in the set {0, 1, 2, ...,9) we should have N(i,n)/n tending to 1/10 as n tends to infinity. But a specific 2-digit number, say 47, should occur among all two digit blocks about 1/100 of the time etc. Borel called a number normal (in base 10) if every k-digit number occurred among all the k digit blocks about 1/10k of the time.

> *He called a number absolutely normal if it was normal in every base b.*
>
> *Now Borel was able to prove that, in one sense, almost every real number was normal. His proof of this involved showing that the non-normal numbers formed a subset of the reals of measure zero. There were still an uncountable number of non-normal numbers, however, which was easily seen by taking the subset of all real numbers with no digit equal to 1. These are uncountable as can be seen using Cantor's diagonal argument, but clearly they are all non-normal. Clearly no rational is normal since eventually it ends in a repeating pattern. However despite proving these facts, Borel couldn't show that any specific number was absolutely normal. This was achieved first by Sierpinski in 1917.*

This is actually our first contact with a fact from measure theory and that is a good motivation.

> *It is reasonable to ask whether π, √2, e etc are normal. The answer is that despite "knowing" that such numbers must be absolutely normal, no proof of this has yet been found. In fact although no irrational algebraic number has yet been proved to be absolutely normal nevertheless it was conjectured in 2001 that this is the case.*

Borel's accessible numbers nicely mirror some of my own (older) thoughts on this subject:

> *An accessible number, to Borel, is a number which can be described as a mathematical object. The problem is that we can only use some finite process to describe a real number so only such numbers are accessible. We can describe rationals easily enough, for example either as, say, one-seventh or by specifying the repeating decimal expansion 142857. Hence rationals are accessible. We can specify Liouville's transcendental number easily enough as having a 1 in place n! and 0 elsewhere. Provided we have some finite way of specifying the n-th term in a Cauchy sequence of rationals we have a finite description of the resulting real number. However, as Borel pointed out, there are a countable number of such descriptions. Hence, as Chaitin writes in [6]:-*
>
> *Pick a real at random, and the probability is zero that it's accessible - the probability is zero that it will ever be accessible to us as an individual mathematical object.*

We now reach 'the computer'. Note the obvious similarity to accessible numbers (memory peg), the 'only' difference being the precision with which we can compute a number compared to describing it in English.

> *In 1936 Turing published a paper called On computable numbers. Rather than look at the real numbers which can be described in English, Turing looked at a very precise description of a number, namely one which can be output digit by digit by a computer. He then took Richard's paradox and ran through it again, this time with computable numbers. Clearly computer programs, being composed from a finite number of symbols, are countable. Hence computable numbers are countable. List all computer programs -- in fact they will all occur in the number c above. Create a new real number t by Cantor's diagonal argument whose n-th digit is defined as follows. If the n-th block is a program which outputs a real number, make the n-th digit of t different from the n-th digit of the computable number which is output. If the n-th block is not a valid programme to output a real number, then make the n-th digit of t equal to 1. Now t cannot be*

computable since, by construction, it differs from each computable number in at least one digit. However, we have just given a recipe to produce the number which could easily be programmed to output t, so t is computable.

Although the "English descriptions" of Richard's paradox must hold the key to the paradox, in this case our "computable numbers" are very precise and not subject to the same difficulties. Do we really have the ultimate paradox which shows that the real numbers are inconsistent? No! So where is the error in our paradox? The error lies in the fact that when we run the computer programmes we do not know whether they will ever output an n-th digit. We can deduce from this argument that it is impossible to tell whether a computer programme which has output k digits will ever output a k+1-st digit.

Note the subtle difference relative to Richard’s paradox. It is one that actually sheds a new light on the language version. Unlike Richard’s paradox, where it is legitimate to claim that we described a new number in a finite number of words, the recipe is one for a program that would run forever, but the program is nevertheless easily programmed in a finite number of statements. Nevertheless, the paragraph is too short to gain new insights into the halting problem (a problem that still escapes me). I also doubt the paragraph’s conclusion, but I leave this for another day.

Finally, note that many if not all of these mathematicians are one have found interesting during our studies, and they all took this issue very seriously, up until Turing. What an ending for this journey, tying to computing.

References

1. David E. Joyce, ‘Notes on Richard Dedekind’s Was sind und was sollen die Zahlen?’.
2. http://www-history.mcs.st-and.ac.uk/~john/analysis/Lectures/A3.htm
3. I.M James, ‘History of Topology’.
4. http://www-history.mcs.st-andrews.ac.uk/HistTopics/Real_numbers_1.html, 2 and 3. We have a pdf version in the library.
5. Cornelius Lanczos, The Variational Principles of Mechanics.
6. http://en.wikipedia.org/wiki/Richard%27s_paradox
7. Timothy Gowers ed. ‘The Princeton Companion to Mathematics’.
8. Richard Dedekind. Essays on the Theory of Numbers. I. Continuity and Irrational Numbers II. The Nature and Meaning of Numbers. Translated by Wooster Woodruff Beman.
9. Michael Spivak. Calculus.
10. Solomon Feferman, Conceptions of the Continuum.
11. Ernst Hairer. Analysis By its History.
12. E. Landau. Grundlagen der Analysis.
13. Morris Kline. Mathematical Thought From Ancient to Modern Times.

.