The Concept of Real Numbers

Numbers are abstractions intended to capture some quantitative properties of physical objects and to make it possible to compute the answers to various questions. Different questions require different kinds of numbers to be useful for answering them. We began historically, as a race, with the simplest kinds of numbers, positive integers, for the simplest counting problems, and invented new numbers as new problems arose. The "real" numbers are designed to solve the problem of measuring lengths of line segments. (There is nothing any more real about them than there is about any other kind of numbers. All numbers are imaginary constructs. In particular, they are no more real than are the so-called "imaginary" numbers which we probably won't have time to study.) Measuring lengths is not such a simple problem and consequently the real numbers are rather sophisticated things. (The difference between rational numbers and real numbers is much greater and subtler than the difference between real numbers and imaginary numbers.)

A closely related concept to that of a number is the concept of a "numeral", which is a symbol that is used to represent or to "name" a number. A number is more of an idea, whereas a numeral is more of a physical object, something you write on paper for instance. One of the difficulties in conceiving of the set of all real numbers is the problem of finding names or numerals for all of them. There are just too many real numbers for us to be able to easily give names to all of them.

To understand this a little better, think about a simpler collection, the set of all positive integers, the counting numbers 1,2,3,4,5,6,7,8,9,10,11,....... Notice how I had to eventually give up on writing them all down, and just put dots? None of us has ever actually seen all the numerals given to even this fairly primitive collection, the set of counting numbers. How do you know they all exist? That is, how do you know the set of positive integers as a whole, exists? It would be an infinite collection, and there is, in our physical world, no infinite collection of anything. If you think about it, not only the grains of sand on all the beaches and oceans are finite in number but even the total number of electrons in the universe is finite. So having started from a pile of rocks, like the cyclops in Ulysses enumerating his sheep, we have arrived, (some of us anyway), at an act of faith by which we assume the existence of an infinite collection of things called the counting numbers, which we assume to have the familiar arithmetic properties we know to hold for the few relatively small numbers we actually have used in our lives.

We use the symbol Z^+ to denote the collection of all the counting numbers. [Z is the first letter of the German word for "number" (and for "to count") and "+" of course stands for "positive". Why the German abbreviation? Well, probably it was a German who first made it up, maybe Richard Dedekind, or Leopold Kronecker, two nineteenth century mathematicians who are very famous, at least in the trade. All you budding mathematicians should think about learning German and French as soon as you have a chance. For now, "to count" is "zahlen".]

Now even though the set of positive integers, Z^+, is infinite, we sort of have names for all of them, at least in the sense that we know how to proceed along from one to the next , successively constructing the names of as many counting numbers as we want. That is, starting from a basic finite alphabet of the ten symbols {0,1,2,3,4,5,6,7,8,9} every one of the positive integers has a "name", its numeral (like 3728552758166338), written using only a finite number of these symbols. Just putting in a minus sign we get names also for all the other integers.

It is no harder to represent the rational numbers, those useful numbers that provide solutions to equations like $bx-a = 0$, where a, b are integers, and $b\neq 0$. I.e. just define a rational number to be given by a pair of integers (a,b) where $b\neq 0$, but we agree that the two pairs (a,b) and (c,d) shall represent the same rational number if (and only if) $ad-bc=0$. We also write a/b, of course, for the number represented by (a,b). So to name a rational number you just need two integers and thus a finite number of the basic digits. (The difference between a number and a numeral is pointed up by the fact that 2/3 and 4/6 for example are two different numerals (or names) for the same rational number.)

Now obviously in the naive sense there are lots more rational numbers than there are integers, since the integers correspond to just those rational numbers of form (a,1), i.e. those whose second part (denominator) is one. But in another sense there are in fact no more rational numbers than there are integers! What we mean by that is that you can "count" off the whole set of rational numbers just using the integers. How? Let's restrict to positive ones for simplicity. We are just trying to give the idea anyway. Okay, here is a complete list of all positive rational numbers: {1/1, 1/2, 2/1, 1/3, 3/1, 1/4, 4/1, 2/3, 3/2, 1/5, 5/1, 1/6, 6/1, 2/5, 5/2, 3/4, 4/3, 1/7, 7/1, 3/5, 5/3, 1/8, 8/1, 2/7, 7/2, 4/5, 5/4, ...}. Get the idea? Just start with all those whose numerator and denominator add up to 2, then those whose num. and denom. add up to 3, then those whose num. and denom. add up to 4, and so on, but don't bother to put in a fraction that has already occurred earlier, such as 2/4, which occurred as 1/2. Now just start at the beginning of this list and number them off using just the positive integers: {1/1 (#1), 1/2 (#2), 2/1 (#3), 1/3 (#4), 3/1 (#5) , ...}. Now obviously there are enough positive integers to rename each rational number this way , i.e. 3/1 could be called just 5, and 1/3 could be called 4. Of course this is confusing and you would need a dictionary now to keep up with what rational number is meant when you see a positive integer like 26. So we would never want to do this in practice but the point is that we could if we wanted, i.e. there are enough positive integers around to make it possible.

That was my point: there are just as many positive integers as there are rational numbers . In fact I claim that if there were not, then it would not have been possible to represent the rational numbers with symbols (like 164/789) in the usual way, using just the basic numerals {0123456789}. That is to say, any set of numbers at all to which one can assign (finite) names spelled with symbols from some finite alphabet, is no more numerous than the positive integers. To see that, suppose we have invented any collection of numbers at all, say the Paideia numbers, and we denote them by P, for example. Moreover suppose it is possible to give names (i.e. numerals) to all of them starting from some finite alphabet of basic symbols, and assume that each one of our numbers has a name consisting of only a finite number of our symbols, like a word spelled with a finite number of letters. Then I claim that the set of Paideia numbers has no more numbers in it than does the set of positive integers. To see that, just line them up sort of like the way we did with rational numbers: i.e. first write down all the one-symbol numerals (one-letter words). There can only be as many of those as there are letters or symbols in our alphabet, right? Okay, next write down all the two-symbol numerals (two letter words). If you think hard you will see there are only as many of those as the square of the number of basic symbols, (i.e. how many two-digit numerals are there using just the usual basic ten digits? Don't forget 00.) Okay, continue this way and you have an infinite list in which every numeral eventually appears. Then you just start at the beginning and number them off, using the positive integers as before. As before, you have enough positive integers to number everything in the list, and that means that there are as many positive integers as there are objects in the list, i.e. as many positive integers as Paideia numbers!

Now you may think, after those arguments (if you believe them), that actually <u>no</u> set of things can have more elements in it than does the set of positive integers, but interestingly, that is not true!! It will turn out that there are actually many, many, more <u>real</u> numbers than positive integers, in every reasonably subtle sense of the word "more". Indeed, as we said in the beginning, one big problem with understanding real numbers is that there are so many of them that there is no way to give (finite) names to all of them, starting with just a finite set of basic symbols, like {0,1,2,3,4,5,6,7,8,9,}. That is why most of us don't quite believe there are so many real numbers as there really are: we haven't seen their names in the newspaper or in our books so they must not exist! Well, that isn't so nice, so we <u>will</u> name them all, and using just the ten basic symbols too, but we will have to give up something, so we will give up the idea that the name should be finite. In terms of our analogy with letters and words, we will use a finite alphabet, but we will allow words to have an infinite number of letters.

We have already met the phenomenon of using an infinite numeral to represent a single number when we tried to give the decimal representation of a simple fraction like 1/3 = .3333333..........., where the 3's on the right continue forever. Of course this means that decimals are sometimes <u>worse</u> for representing rational numbers than were the good old pairs of integers. But the positive side to the story, I claim, is that decimals allow us to represent many more numbers than just the rational numbers: i.e. I claim that <u>every</u> real number, rational and irrational alike, can be represented by an infinite decimal. (The finite decimals also occur among the infinite decimals, as those which are zero forevermore after a certain point in the expansion).

Okay, so what <u>is</u> a real number, why do they correspond to infinite decimals, and why are there a lot more of them than there are rational numbers? First a real number is a number that corresponds to a length, i.e. a number that names the ratio between two line segments, (the length of a segment is the ratio between it and the" unit " segment). So if we start from the points on the "real line" , i.e . if we start with a line and mark off a point to be the zero point, and then mark off a point to the right of zero to be the unit point, then there ought to be exactly one real number for each point of the line. So the real numbers ought to be some collection of numbers that correspond exactly to the points on the line. So to justify our claim that infinite decimals are a reasonable way to represent real numbers, let us try to show that there is a natural correspondence between infinite decimals and points on the line. I.e. Choose a point x, and assume for simplicity that it lies to the right of zero. Then lay off copies of the unit interval end to end, on the line,starting from zero, until you get one whose right end point does not go to the right of x , but so that the right end of the next unit interval does go to the right of x. The number of unit intervals you have laid off, i.e. the largest number that do not reach actually to the right of x, is the integer part of the decimal we are constructing. Now of course we think it is obvious that this procedure can be carried out, but in fact in doing so we have tacitly assumed that the real line has a special property which should be made clear, the

Archimedean property: Given any point x on the line, and any positive length (determined as the interval between two distinct points), if you lay off enough copies of the given length, starting from zero on the line, you will eventually (in a finite number of steps) go past x.

(Saying this plainly is really not a waste of time; i.e. since the line really only exists in our minds, and all our minds are different, it is prudent to spell out any property which we will take to be true about it, instead of just assuming that everyone is thinking the same thing. So this is an axiom about the line, and its honorary name implies that somebody was thinking rather deeply about this question a long time ago.)

Now that we have the integer part of the decimal, we get the tenths part in a similar way; i.e. subdivide the unit length into ten equal parts, and then take one of these tenths and start laying off copies of it end to end starting from the point marking the integer part of x. Again, using the Archimedean property, there will be a segment which does not itself reach to the right of x but such that the next segment will do so. The greatest number of segments that do not reach to the right of x is a number between zero and nine, called the tenths digit in the decimal expansion of x. We get the hundredths digit the same way, and continuing, we can construct as much of the infinite decimal as we want.

You might want to think of an infinitely finely divided meter stick, on which you are measuring off a length. First you read off the number of meters, being careful not to go over the actual length you are measuring, i.e. read off the largest number of meters that is not greater than your length. Then read off the decimeters, again the largest number that is not greater than your length. Then focus your eyes finer and read off the millimeters. You now have a number like 4.95, which means that the actual length is at least equal to 4.95 meters but is less than 4.96 meters. Since the meter stick is infinitely finely scored you can continue as long as you like, getting ten or a hundred, or a billion decimal places of accuracy, if you wish.

Even though we cannot construct the entire infinite decimal in a finite amount of time this way, recall that neither can we actually construct the whole of the familiar collection of positive integers either, (remember how we had to just put dots eventually when we were trying to write down all the positive integers?). Thus, just as we know how to count up the list of positive integers as far as we may wish, so also we are able in this way to construct as many digits of the infinite decimal expansion of x as we wish. In this situation, most mathematicians are content to say that we have constructed, or at least that we have given a prescription for constructing, the full decimal expansion of x, and to leave matters there.

<u>To summarize</u>: if we assume the plausible statement that the line has the Archimedean property, then each point x on the line gives rise to a specific infinite decimal expansion, which we may think of as (an infinite numeral representing) the real number which corresponds to x.

Now what of the opposite correspondence? That is, suppose we start from an arbitrary infinite decimal; is it necessarily the real number coming from some point on the line ? The answer is essentially yes, with one class of exceptions: Certain points on the real line, the ones having finite decimal expansions, also have another different expansion which is infinite. I.e. a number like 14. 26000000000........, has another expansion namely 14. 2599999999...... However, the method that we described of assigning a decimal to a point on the line will not allow you to come up with the second of these two expansions. Thus you could say that , by our constuction, the points of the real line correspond exactly to those infinite decimals that never become eventually equal to all nines. It will be better I think though, if we go ahead and think of real numbers as given by all infinite decimals and just remember that every infinite decimal that ends in all zeros (i.e. every "finite" decimal) can be written again, in a different way, as an infinite decimal that ends in all nines. This points up the fact that the decimals are numerals , i.e. merely names for the numbers and not the actual numbers themselves. Some real numbers have more than one name, that's all. (This was also true of fractions, each of which has many names, 1/2 = 3/6, etc.) Our constuction of an infinite decimal starting from a point on the line was designed to pick out, in those cases where two names are possible, the name ending in all zeros.

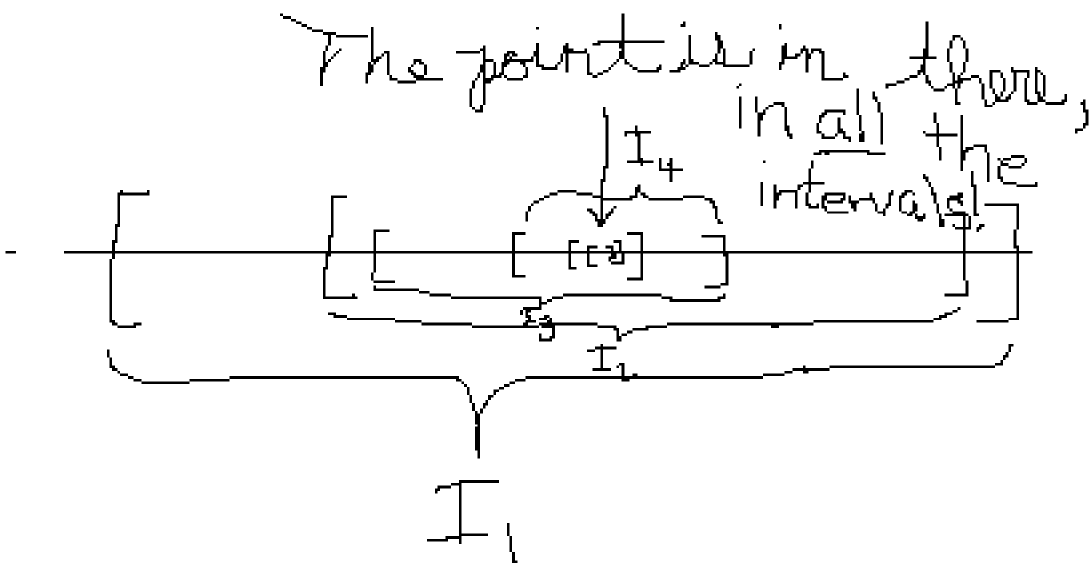
Okay, so let us now start from any infinite decimal at all, possibly even one that ends in all nines, and explain how to construct the corresponding point on the real line. We just give an example: say the decimal is given by 19.1911911119111119111119........., you get the idea, add an extra one each time before the next nine. This infinite decimal should really be thought of as an infinite sequence of finite decimals : 19, 19.1, 19.19, 19.191, 19.1911, 19.19119, 19.191191, ... , and so on. To this sequence of finite decimals we associate the following infinite sequence of closed bounded intervals on the real line for which these decimals are the left end points: [19,20], [19.1, 19.2], [19.19, 19.20], [19.191, 19.192], [19.1911, 19.1912],........, and so on. Note that this is a "nested" sequence of intervals, in the sense that each interval completely contains the next interval in the list. I.e. any point that lies in any one of these intervals, automatically lies also in every previous interval in the list. Note also that the length of succeeding intervals is growing shorter: the first interval has length one, the second has length 1/10, the third has length 1/100, and so on. <u>Then we assert that there is precisely one point x on the line, that lies in every one of these intervals, and that x is the point corresponding to the number we started with.</u>

If you look closely at this construction you can check that it reverses the one that assigned a decimal to a point. I.e. if you take this point x and assign a decimal to it in the way we did earlier, using Archimedes' axiom, you will get back exactly the decimal 19.191191119...., that we started with. In general if you start with any infinite decimal and construct a point on the line this way, and then construct a decimal from that point by our earlier construction, you will always get back the decimal you started with, <u>unless you started with a decimal that ended in all nines</u>. In that case you will get back the decimal ending in all zeros that represents the same point. The other direction always works; i.e. if you start with a point, then construct a decimal and then use the decimal to construct a point this way, it will always be the point you started with. You should stop and think about these constructions until you understand that they do indeed reverse each other. Now we must make our second important observation about the real line: the claim that the nested sequence of closed bounded intervals actually does contain a point, so that this construction really works, cannot be proved, but is in fact the deepest and subtlest property of the real line, the one that guarantees the existence of plenty of irrational numbers, the

Completeness property of the real line: Every nested sequence of closed bounded intervals contains at least one common point. I.e., if $I_1, I_2, I_3, I_4, \ldots$ is a sequence of closed, bounded intervals on the line, such that each interval I_n completely contains the next interval I_{n+1}, then there exists at least one point of the line that lies in every one of the intervals.

(To be precise, a closed bounded interval consists of two points of the line together with all the points lying between them.)

Now this is another axiom about the line, which I think is far from obvious, but is one attempt to state precisely what is meant by saying the line is "continuous", i.e. has no gaps or breaks in it.



To see a bit better what this says, let's examine some closely related statements which are false, so that we can try to understand the difference. Note that we have carefully assumed that our nested intervals are closed and bounded. What happens if we forget to assume these things? I claim we get trouble. For, consider the sequence of unbounded intervals $[1,\infty)$, $[2, \infty)$, $[3, \infty)$, $[4, \infty), \ldots\ldots$, where each interval goes all the way to the right "end" of the line. You might possibly think that since these intervals are bigger than bounded intervals that they would be more likely to contain a common point but it is easy to see that there is no point of the line which is in all these intervals. I.e. x belongs to the interval $[n, \infty)$ if and only if x is at least as large as n, and no real number x can be as large as every integer n. (Why not? We assumed it, in the Archimedean axiom!)

Now what happens if we make the intervals bounded all right but forget to make them closed? For example, take the sequence (1,1.1), (1,1.01), (1,1.001), (1,1.0001),.......... This time the intervals are open which means that the endpoints are not included. It is not hard to see that there is no point contained in all these intervals, since any point in all these intervals would have to be bigger than one, in order to be to the right of all the left end points , but I claim (you really need the Archimedean axiom to prove it) that any number greater than one would eventually be too big to stay to the left of all those right endpoints. So this nested sequence of bounded open intervals does not have any point common to all of them. Still think that the axiom was so obvious? Well, anyway we are going to assume it.

Now we are going to show, using the completeness axiom, that each infinite decimal gives us a unique point on the line, and since this axiom talks only about existence we need to say something about uniqueness. Fortunately, this is a corollary of the Archimedean property that we already have. Let's see what we mean by that. Suppose you pick a point x to the right of zero on the line. We already know that if you start at zero and lay off enough copies of the unit iterval or any interval bigger than one point, that you will eventually go past x. That says that x cannot be "infinitely far away from zero". But what about the possibility that x is infinitely close to zero? I.e. we claim:

Archimedean property II (Density of rational points): If x is any point to the right of zero, and if we consider any other finite interval extending to the right from zero, such as the unit interval, then it is possible to subdivide that interval so that the first subdivision will occur between zero and x. (In terms of numbers this means that if x is positive and n is big enough then 1/n is less than x. Of course to prove it algebraically you just take n greater than 1/x.)

To see this geometrically, assume first that we are dealing with the unit interval, just so the interval will have a name. Then consider the case where the end of the unit interval, the unit point, is already to the left of x. Then there is nothing to do, and we are finished. So assume now that the unit point, call it 1, is to the right of x. Now use the Archimedean property to find n so large that n copies of the interval from zero to x, whose right endpoint we will call nx, reaches past 1. Now just subdivide both intervals [0,nx] and [0,1] into n equal pieces. Since nx is to the right of 1, the first subdivision of [0,nx], which is x, should be to the right of the first subdivision of [0,1], which is what we wanted to prove. Now that word "should" is in there because I don't see immediately how to prove that from Euclidean geometry, so either work it out yourself or assume Arch.II as another axiom if you want to. (Fine points like this on whether one point is or is not between two other points are usually not discussed too carefully in elementary geometry courses, so it is hard to remember whether this is proven in Euclidean geometry or not.)

Let's assume the following form of the axiom which is actually equivalent again to the previous two versions:

Archimedean axiom: Let x and y be any two different points at all, equal to or to the right of zero on the real line and let 1 be any point at all to the right of zero, which we call the unit point. Then it is possible to choose a positive integer n so large that, if we subdivide the unit interval into n equal pieces, and begin laying off copies of the first piece end to end, going to the right from zero, then at least one of the end points will fall between x and y. (This says x and y, if different, cannot be infinitely close to each other, as measured by our unit interval.)

Corollary: The sequence of intervals above associated to the decimal 19.191191119......, do indeed contain exactly one point.
proof: That they contain at least one point is guaranteed by the completeness axiom so we must show that they do not contain more than one point. It suffices to check that if x and y are any two different points of the line, that they cannot both lie in all the intervals. Well if x and y both lie in the first interval [19,20] it means that no integer point lies between x and y. OK. But if also both x and y lie in the second interval [19.1,19.2] it means that no tenths point lies between x and y either. If x and y were both in the next interval as well it would mean that no hundredths point lies between x and y either. Well, the Archimedean II axiom says eventually one of those subdivision points has to lie between x and y, which means that only one of them , x or y but not both, lies in that interval. Hence two different points x and y cannot both lie in all the intervals.

We have proven reasonably carefully, I think, (and you may well think we have reduced it to ashes and raked the residue with a fine tooth comb) the following theorem:

Theorem: Assuming that the real line satisfies both the Archimedean axioms and the completeness axiom, the choice of a zero point and a (different) unit point on the line establishes a natural one to one correspondence between the points of the line and those infinite decimals which do not eventually end in all nines.

Definition: We define the set of real numbers to be the set of all infinite decimals, with the condition that a (positive) infinite decimal ending in all nines shall be considered as representing the same real number as the decimal which begins with the same digits up to the last one before the nines begin, but in which that last digit (which is less than nine) is increased by one and is then followed by all zeros.

Well we have defined the real numbers as a set of numerals, which is a grand and wonderful accomplishment. There are indeed enough of them now I claim, although we have not proved it yet, to provide square roots of any positive number you like, something the rational numbers did not provide as we know. So we should pause briefly to celebrate.
Not for too long however, since numbers are supposed to admit of being used for calculations after all, and as we remarked in class it is not at all obvious even how to add two infinite decimals! Since if you think about it, this is the assertion that for any two numbers there exists a number that deserves to be called their sum, it is not too surprising that a proof will use the

only existence property that we have, the completeness axiom. It will be useful for us to develop some of the other equivalent forms of the completeness axiom. One which is rather popular is a form that actually contains both the completeness axiom and the Archimedean axiom in one statement! That is the famous "least upper bound" property, which we will discuss in a moment.

Right now we pause to show that there really are **more** real numbers than there are positive integers, in a rather precise sense. Since the positive integers are the counting numbers, we shall discuss which sets can be counted by them, and then we shall call such sets "countable". Now usually we think, I would guess, that counting a set of things means to assign a number to the whole collection that tells how many there are. But if you think about the process of counting for a second, you will recall that it really proceeds by assigning one number to each of the elements of the collection, beginning with assigning the number one to the first element, and then the number of elements in the collection is said to be equal to the number that gets assigned to the last element.

We want to enlarge on this a little to say what it means to "count" an infinite collection. It just means that you start counting off the elements, putting them in some convenient order, assigning one to the first element, two to the second, and so on, and continuing on through the whole collection. If the collection is infinite, so that there is no last element, we will still say that we have counted the elements provided everyone of them gets counted off eventually. One way to think of it is to make a numbered list like this but which goes on forever:

1.________________
2.__________________
3.__________________
4._________________
5.____________________
....

The list should be thought of as having an infinite number of blanks for you to fill in with the things you are trying to count. If it is possible to figure out some way to order your set of things so there is a first one, then a second one, then a third one, and so on, then you really can put your whole collection of things down on this list so that they all get counted. If this can be done we call the set "countable", even if it may be infinite. So the real problem, for counting infinite collections is to find the right way to order them. For instance when we counted off the positive rational numbers earlier, we took them completely out of their natural ordering by size, since we all have noticed that in that familiar ordering there really is no first, second or any such thing. So the whole difficulty is in being clever enough to dream up a good way to reorder a set that you are trying to prove to be countable.

Now here is a general principle for recognizing countable infinite sets: Suppose that S is a colection of things which can be broken up into a countable number of subcollections, S_1, S_2, S_3, S_4, ... and so on, each having a finite number of elements . Then S is a countable set. To see this, just start by putting the elements of S_1 into the list using up as many spaces as needed

. Then follow them in the list by the elements of S_2. Keep on like this and clearly each element of each set will eventually get put on the list. This is how we showed the positive rationals to be countable, by breaking the set of all of them up into subsets according to how large is the sum of the numerator and denominator. That is, we put S_1= {all positive rationals with num. plus denom. equalling 1}(there are none of these, but so what); then S_2={all positive rationals with num. plus denom. equal to 2}; and so on. This is also how we showed the "Paideia" numbers are countable. I.e. we let the first set be the set of all numerals with one digit; the second set contained all numerals with two digits, and so on. Since each of these sets is itself finite we can make our argument as before. Ok, now if you have understood that, here is a challenging little problem for you:

Definition: A number which is a solution of some polynomial equation $a_nx^n + a_{n-1}x^{n-1} +..... + a_1x + a_0 = 0$, with integer coefficients, is called an algebraic number.

(For example, $2^{1/2}$ is an algebraic number because it is a solution of the equation $x^2-2 = 0$, which has integer coefficients.)

Problem: Prove that the set of all algebraic numbers is a countable set. (Think about it for a bit , and if you need a hint, look at the bottom of the next page.)

Now we at last try to dispel the perfectly reasonable but incorrect idea that all sets are countable:

Theorem: (Georg Cantor) The set of real numbers is not countable.

Proof: We will just show that no list of real numbers can contain all of them. Indeed we will restrict ourselves just to those real numbers lying between 0 and 1 and whose decimal expansion contains only zeros and ones. What we will do is assume that we are given a list of such reals and then give a way of cooking up another such real that cannot possibly be in the list. That will show that the list cannot be complete.

So let us take an example of one possible list:

1) .000000000.....
2) .1000000000...
3) .11000000000...
4) .111000000000...
5) .1111000000000...
....

Now of course it is easy to see that .0101010101010101... for example is nowhere in this list, but let's try to construct another number, not in the list, in a more systematic way, a way that has some hope of working on all other lists as well. Notice first of all that in order for two decimals of this type to be different they only need to be different in one digit. So for instance, to construct a decimal that is different from the first one in this list, we only need to let it begin with a 1, instead of a 0. So we let our number start out as .1 , and then we have already made it different from the first number in our list, and we are nowhere near finished constructing it. How should we continue it so that it will be different also from the second number in the list? Since we are ready to choose the second digit, and since the second number in the list has 0 as its second digit, we only need to choose our second digit as 1. So our decimal now looks like .11 , so far. Now look at the third digit of the third number in the list, which is again a 0, and conclude therefore that we should choose a 1, and we have .111, for our number so far. Get the idea? We are led to the number .111111.... which is different from every number in the list. Now let's try it in general.

Suppose we are given some infinite list of infinite decimals like those above:
1) $.a_1a_2a_3a_4$.........
2) $.b_1b_2b_3b_4b_5$........
3) $.c_1c_2c_3$...........
....

Now to construct an infinite decimal which is not in the list, we do as we just did. To figure out a way to write it, let's make a rule. As we know, every digit in every number in the list is either a 0 or a 1. Let us put a " ’ " symbol over a digit to change that digit into the opposite choice of digit, 0 or 1. Thus if a is 0 then a’ is 1, and vice versa. Then our number which we claim is nowhere in the list, can be written as $.a'_1b'_2c'_3d'_4$.....
That is, our number has a different first digit from the first decimal in the list, so it does not equal the first decimal. It also has a different second digit from the second decimal in the list, so it does not equal that second decimal either. Indeed our number cannot equal any decimal in the list since it differs from each decimal in at least one digit. Thus we have constructed a real number whose decimal expansion has only 0's and 1's and which is not in the given list. Thus no one list can contain all real numbers of that type. That is, the real numbers of that type are "uncountable". This implies also that the "larger" set of all real numbers is uncountable too, of course.

Now think back to the problem about proving that the set of all algebraic numbers <u>is</u> countable. Here's the hint if you want it: Let's say that for a given equation with integer coefficients that we will call the "size" of the equation the integer you get by adding together the degree of the equation and the absolute values of all the coefficients. For example, the size of $2x^3 + 17x - 5 = 0$, is equal to 3+2+17+5=27. Then let the set S_1 be the set of all equations of size 1, and S_2 be the set of all equations of size 2, and so on. Then I claim that there are only a finite number of equations in each set. See if you can prove that. Then I claim also that a given equation has only a finite number of solutions. You may take that for granted for now but we will prove it later as a consequence of the "factor theorem" for polynomials. Do you see how to finish the problem now? Stop and think about it for a minute or two. [If you don't get it read on: Now we can make a list of all algebraic numbers by taking all the solutions of equations in the set S_1

and putting them in the list first, one after another until we have all of them. Then after those, we can put down all the solutions to the equations in the set S_2, and so on. In this way we will eventually list all solutions of every equation, i.e. we will list every algebraic number, so that we have proven that the algebraic numbers are countable.] Now we get an interesting bonus! Since the algebraic numbers are merely countably infinite, whereas there is a much larger, uncountable infinity of all real numbers, it follows that some real numbers, indeed, most of them, are not algebraic!