

1. Counting: A silver platter from India yet unrecognized

When children learn to count, they accomplish two hidden tasks. One is the visualization of the infinite sequence that we call the natural numbers: 1, 2, 3, 4, . . . The other is learning the particular rules in their language, that generate them from finite many words. If the number system is stupid, like the Roman numerals, then finite many symbols are not really sufficient, but the infinity is still conceived. And yet, the Roman soldiers were unable to collect taxes properly and this could have been helped by the Arabic numbers. Later, one pope still tried to defy the new numbers and regarded the zero as the devil's work. The truth is that the Arabs brought the system from India, where they calculated with millions, way back in time. Beside the Indian origin, the other less known fact about the Arabic numbers is that like many other eastern influences, they penetrated Europe through the other wing of the church in Byzantium or Constantinople. The slaves turning into peasants meant a mass of people who had to count, so the practical superiority of the Arabic number system became unavoidable.

But at the level of counting, all this is only in the background.

The best practice of counting as a deeper knowledge, is the exercise of counting in increments.

For example, starting from 20, the child can go in steps of 3, as 23, 26, 29, 32, . . .

This is what we later call an arithmetical sequence, with the 20 start and 3 increment. If the start is itself the increment, then we can simply call the sequence as the multiples of the start, like at 7, 14, 21, 28, . . . Another way to say that something is multiple of 7, is that it is dividable by 7, but this is not quite the same because the start 7 is also dividable by 7.

So in abstract notation the multiples are $m \cdot n$ for any n starting number except 1 and for any m multiplier except again 1. The starter of course can be obtained as $n = 1 \cdot n$ that is with the trivial 1 multiplier too. The exchangeability of multiplication, that is $m \cdot n = n \cdot m$ means that any multiple appears not only as m multiple of the n start but also as n multiple of m .

This exchangeability is reflected in an other name for the multiples, called as composites. Indeed, being $m \cdot n$ can be said as being composed from m and n .

The non composites, that is numbers that can not appear as multiples are called primes.

Strangely, the healthy vision of the multiple sequences is working against the concept of compositeness and primeness. The first visual mistake is that the starter melts into the sequence, so we have to emphasize that only the later members are the composites. But the second mistake is then to think that all starters are non composite or prime, which is absurd because any number can be a starter. So the best is to look at all the possible sequences:

2 , 4 , 6 , 8 , . . .

3 , 6 , 9 , 12 , . . .

4 , 8 , 12 , 16 , . . .

5 , 10 , 15 , 20 , . . .

A starter can definitely not appear later, that is neither in its own sequence or underneath because only bigger numbers appear there. So a starter is prime simply if it didn't appear earlier. The first starter 2 is then automatically a prime. The next 3 is again because before only the even numbers appeared. But then the starter 4 is not a prime because it appeared in the first sequence as multiple of 2. The next starter 5 is again a prime and then the 6 is not because it already appeared. In fact in two sequences as multiple of 2 and 3.

These multiple sequences are the correct way to introduce the primes but usually in schools it's not done this way, rather merely defining them as non composites, because the composites themselves seem so obvious. Like $4 = 2 \cdot 2$ tells at once everything without multiples.

Instead of the infinite multiple sequences, a quite different direction of abstraction is pursued. Namely, the reverse of the multiple as dividability. We develop a feel for dividabilities but if we are asked what it really means that a k is dividable by an n then we merely go back to multiplication: $k = m \cdot n$. The word "divider" includes two trivial ones for any number k . Namely, 1 and k itself. They together give the trivial product form of any number: $k = 1 \cdot k$. So, to exclude these trivial products, it's enough to exclude one of them from the dividers.

Since the 1 is the same for all numbers, it's logical to exclude this. So, all other, that is non 1 dividers of a k are called the factors of k . Thus, we allow k itself as factor if k is not 1. This may seem illogical now but will be perfectly useful.

Now, the exact definitions go as follows:

Composites are numbers that are non trivial products, that is products without the use of 1.

Thus, the number 1 is obviously not composite. The other non composites are the primes.

So we have three class of numbers. The trivial unit number 1, the composites, and the primes.

The $k = m \cdot n$ composites are usually broken down into further factors by writing the m and n factors into non trivial products themselves. In fact, we continue this as far as possible. Clearly, the last factors then must be primes, so the initial k composite is broken down into primes.

The uniqueness of these final prime factors, regardless how we did our factorization, is the most important role of the primes. If we allowed 1 as prime, that is as factor, then such uniqueness would be obviously impossible because we could always include new 1 "factors".

2. Factorial: A tricky number that helps a lot

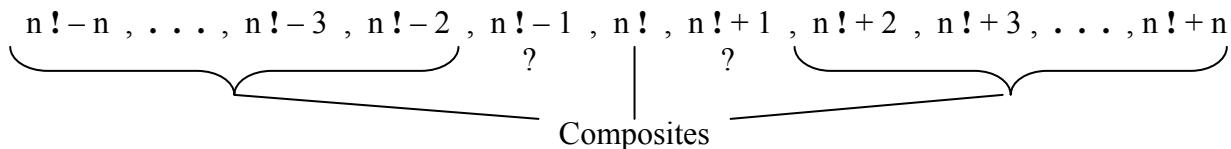
Multiplying numbers is not a condition of seeing the multiples, so the child doesn't have to be a good multiplier to go way above his calculation. In fact, it's an advantage to be bad. A concept that can prove this, is the factorial, denoted with an exclamation mark. For example,

$100! = 2 \cdot 3 \cdot 4 \cdot \dots \cdot 99 \cdot 100$ This is obviously a number so huge, that nobody can give it exactly, and yet everybody should explore it. The weird thing about $100!$ is that it is multiple of all the 2, 3, 4, . . . , 99, 100 numbers. Then of course, $100! + 2$ is a multiple of 2, in other words even, $100! + 3$ is a 3 multiple, that is triplet, and so on.

So, $100! + 2, 100! + 3, 100! + 4, \dots, 100! + 99, 100! + 100$ are all composite numbers.

Of course, 100 can be any other n , so $n! + 2, n! + 3, \dots, n! + (n - 1), n! + n$ are always a sequence of consecutive numbers that are all composite. So arbitrary big n gives arbitrary big such block without a prime. This in itself is a great result, but lets observe that instead of adding the 2, 3, . . . , n numbers, we could subtract them, that is count back, so:

$n! - 2, n! - 3, \dots, n! - (n - 1), n! - n$ are always composites too. So actually, we have two twin blocks without prime, and in between them are three numbers $n! - 1, n!, n! + 1$.



The $n! - 1$ and $n! + 1$ numbers could be called the factorial twins. They are definitely not multiples of any of the 2, 3, . . . , n numbers because $n!$ is and 1 isn't. The question marks under them show that we don't know whether they are composites or primes. Indeed, just because they are not dividable by 2, 3, . . . , n they could have factors all bigger than n . If that's the case, then the two blocks combine into a single, so from $n! - n$ up to $n! + n$, all numbers are composite, that is, there are no primes.

Is it possible that the primes would die out completely and after some n number there were no primes at all? The factorial itself can help to disprove this. Namely the factorial twins.

The $n! - 1$ and $n! + 1$ values themselves don't guarantee primes but lets remember, that they are definitely, not dividable by 2, 3, . . . , n .

So now we simply regard the s smallest factor of one of them say $n! - 1$.

This can not be a composite because then its smaller factors were smaller factor of $n! - 1$ too.

So this s smallest factor of $n! - 1$ is a prime.

But it can't be any of 2, 3, . . . , n , so this s is a bigger than n prime.

Euclid showed this proof for the infinity of primes more than 2000 years ago, without Arabic numbers. We don't teach it today, when it should be even more important to avoid the misuse of the silver platter.

3. Prime factorization: The universal multiplier

The basic fact in Euclid's proof above, besides the tricky choice of $n! - 1$, was that the smallest factor of a number is always a prime. So, if $\text{minfact}(n)$ denotes the smallest or minimal factor of n , then we can express n itself as: $n = \text{minfact}(n) \frac{n}{\text{minfact}(n)} = p_1 \frac{n}{p_1}$.

If $\text{minfact}(n) = n$ then of course, n is a prime and $\frac{n}{\text{minfact}(n)} = 1$ can be ignored.

If not then $\frac{n}{p_1}$ can be checked again for its minfact p_2 . And so on repeatedly.

Thus, $n = p_1 p_2 p_3 \dots p_m$. Observe that $p_1 \leq p_2 \leq \dots \leq p_m$.

Indeed, if a smaller than p_i minimal factor would appear after p_i , then it would be minimal factor earlier already.

Equality is possible because the smallest factor can still divide the rest of a number.

For example, $100 = 2 \cdot 50 = 2 \cdot 2 \cdot 25 = 2 \cdot 2 \cdot 5 \cdot 5$

$90 = 2 \cdot 45 = 2 \cdot 3 \cdot 15 = 2 \cdot 3 \cdot 3 \cdot 5$

This minimal factoring process is actually the increasing prime factorization.

An other way to obtain this increasing prime factorization is simply to try the increasing prime numbers as dividers in the decreasing leftover numbers. For example, for 90, the smallest prime 2 is a divider, the result is 45. We try 2 again, but it doesn't divide, so go to the next prime 3. It divides and the result is 15. We try 3 again, and the result is 5. Here 3 doesn't work anymore, so we try the next prime 5.

But this "minfact" function is an excellent theoretical tool too. With it, we can formulate Euclid's proof crystal clearly. Indeed, we only have to realize two facts:

1.) $\text{minfact}(n)$ is always a prime.

2.) $\text{minfact}(n! - 1) > n$.

Thus for every n we obtained a prime above n .

Just as multiplication is a repeat of addition: $3 \cdot 5 = 5 + 5 + 5 = 3 + 3 + 3 + 3 + 3$, we have a shortcut for repeating multiplications too: $5 \cdot 5 \cdot 5 = 5^3 = \text{power} = \text{exponentiation}$.

But just to see how different this is from multiplication, lets observe that now its not true that the two members can change place: $5^3 = 125 \neq 3^5 = 3 \cdot 3 \cdot 3 \cdot 3 \cdot 3$

A very similar feature though is that using 1, we get the number itself: $5 \cdot 1 = 5$, $5^1 = 5$.

On the other hand, the extension to 0 value is different again.

At multiplication, quite logically: $5 \cdot 0 = 0$. But here, we'll have $5^0 = 1$.

The reason is quite simple:

$$\frac{5^7}{5^4} = \frac{\cancel{5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5}}{\cancel{5 \cdot 5 \cdot 5 \cdot 5}} = 5^3 = 5^{7-4} \quad \text{or in a special case, } \frac{5^4}{5^4} = 5^{4-4} = 5^0 = 1$$

This 0 exponent being 1, becomes even more logical in the followings:

Lets use exponents for the repeated prime dividers: $100 = 2 \cdot 2 \cdot 5 \cdot 5 = 2^2 \cdot 5^2$

Since $p^0 = 1$, we can include all primes, even the non really occurring ones:

$100 = 2^2 \cdot 3^0 \cdot 5^2$ Then if another number is produced the same way, like

$15 = 2^0 \cdot 3^1 \cdot 5^1$ we can easily calculate their product with the sum of exponents:

$$100 \cdot 15 = 2^2 \cdot 3^1 \cdot 5^3$$

So the prime factorization of big numbers can be obtained from smaller factors instantly.

4. Under the silver platter: The practical exponents of a base

$$2008 = 2 \cdot 1000 + 0 \cdot 100 + 0 \cdot 10 + 8 \cdot 1 = 2 \cdot 10^3 + 0 \cdot 10^2 + 0 \cdot 10^1 + 8 \cdot 10^0$$

This is the logic of the Indian silver platter, that is presenting numbers from powers of 10. But we could use any other base, say 7.

Indeed, find the biggest 7 power that fits in 2008, which is $7^3 = 343$, namely 5 times.

$5 \cdot 343 = 1715$ and the leftover is $2008 - 1715 = 293$. In this $7^2 = 49$ fits 5 times.

$5 \cdot 49 = 245$ and the leftover is $293 - 245 = 48$. In this, $7^1 = 7$ fits 6 times and leaves 6 remainder, so $2008 = 5566_7 = 5 \cdot 7^3 + 5 \cdot 7^2 + 6 \cdot 7^1 + 6 \cdot 7^0$.

Especially easy is the conversion to base 2, because here the powers either fit in or don't, we don't have to worry about the digits, they are all 1 or 0. The 2 powers are,

$$2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, \dots$$

For 2008, the subtraction sequence is:

$$\begin{array}{r}
 2008 \qquad = \qquad 111101100_2 \\
 \hline
 1024 \\
 984 \\
 \hline
 512 \\
 472 \\
 \hline
 256 \\
 216 \\
 \hline
 128 \\
 88 \\
 \hline
 64 \\
 24 \\
 \hline
 16 \\
 8 \\
 \hline
 8 \\
 0
 \end{array}$$

5. Negativity: Formalizing exponents beyond

The negative whole numbers are very easy to conquer with using the “sea level” as 0.

For example, $2 - 3 + 5 - 7 - 2 + 1 = -4$ and every step is a motion up or down.

The stages are: $\begin{array}{cccccc} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ -1 & +4 & -3 & -5 & -4 \end{array}$

The earlier used logic for exponents can lead to negative ones too.

$$\frac{5^4}{5^7} = \frac{\cancel{5 \cdot 5 \cdot 5 \cdot 5}}{\cancel{5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5}} = \frac{1}{5^3} = 5^{4-7} = 5^{-3}$$

The negative powers of 10 can be used to extended the silver platter from India by the well known decimal point:

$$\begin{aligned}
 2008.305 &= 2 \cdot 10^3 + 0 \cdot 10^2 + 0 \cdot 10^1 + 8 \cdot 10^0 + 3 \cdot 10^{-1} + 0 \cdot 10^{-2} + 5 \cdot 10^{-3} \\
 .305 &= 3 \cdot 10^{-1} + 0 \cdot 10^{-2} + 5 \cdot 10^{-3} = 3 \frac{1}{10} + 0 \frac{1}{100} + 5 \frac{1}{1000} = \frac{305}{1000}
 \end{aligned}$$

This extended silver platter surpasses all the problems, the Greek mathematicians struggled so hard to understand. Especially, because today we not only possess this continuing decimal system, but also, the division process that calculates the actual digits for fractions.

$$\frac{5}{6} = 5 : 6 = 5.000 \dots : 6 = 0.833 \dots$$

$$\begin{array}{r} 48 \\ 20 \\ \hline 18 \\ 20 \end{array}$$

In general, all fractions become $0.\underbrace{b_1 b_2 \dots b_m}_{\text{beginning}} \overline{c_1 c_2 \dots c_n}_{\text{cyclic}} c_1 \dots$

This is obvious because if the divider or denominator was d then in the division process we can only have maximum d different remainders. So, we have to return to an earlier one from which everything repeats. By the way, the reverse problem, that is how to find the fraction for an infinite cyclic decimal is also very simple, namely:

$$0.\underbrace{b_1 b_2 \dots b_m}_{m \text{ 0-s}} \overline{c_1 c_2 \dots c_n}_{n \text{ 9-s } m \text{ 0-s}} = \frac{b_1 b_2 \dots b_m}{10 \dots 0} + \frac{c_1 c_2 \dots c_n}{9 \dots 9 0 \dots 0}$$

$$\text{Example: } 0.2035757 \dots = \frac{203}{1000} + \frac{57}{99000}$$

Now returning to the silver platter vision of the infinite decimals, any $0.d_1 d_2 d_3 \dots$ is actually a point on the $[0, 1]$ interval. Indeed,

$$0.d_1 d_2 d_3 \dots = d_1 \frac{1}{10} + d_2 \frac{1}{100} + d_3 \frac{1}{1000} + \dots$$

So if $[0, 1]$ is a metre, then $0.d_1 d_2 d_3 \dots$ means the point being from the left end at d_1 decimetre + d_2 centimetre + d_3 millimetre + and so on.

Since all $0.d_1 d_2 d_3 \dots$ infinite decimals determine a point on $[0, 1]$ but only the cyclic decimals correspond to $\frac{a}{b}$ fractions, thus we see that most of the points on $[0, 1]$ are not fractions. The fractions are also called rational numbers and so most points are irrational. This was very hard to swallow for the Greeks who were only able to prove with complicated but actual methods that some numbers like $\sqrt{2}$ are not fractions.

6. Under the extended silver platter: Achilles and Anti Achilles paradoxes

The missing vision of the irrationals was actually caused by missing the general fact that the addition of smaller and smaller lengths can be a finite one. Indeed, the so called Achilles paradox is exactly this problem. If Achilles is ten times faster than the slowest runner and $[0, 1]$ the unit interval is a kilometre, then we can create the following competition:

Achilles starts from 0, but gives 500 metre head start to his competitor who thus starts from 0.5. By the time Achilles reaches this point, his competitor will have run one tenth of what Achilles have, that is 0.05. So he will be at 0.55. Then, when Achilles reaches that point, the other will be again 0.005 ahead that is at 0.555. And so on, the competitor will always be ahead, so Achilles can't reach him.

The trick here was to transform the adding up of the infinite many smaller and smaller distances into a problem of time and then, to use the false impression, that if something happens infinite many times, then it is true forever. But this is clearly a nonsense! If the rain starts, at noon, then at infinite many times before, it wasn't raining. Namely, at 11, 11:30, 11:45, 11:50, and so on. We can get times that approach noon or even an earlier time. So the solution of the Achilles paradox is easy. Achilles will reach his competitor exactly at the $0.555555\dots$ point. There are infinite many points and times before this point. Motions go through these infinities continually. So though the distance at which the reach happens is a finite distance, it is made of infinite many smaller ones. Even the feeling that this infinite sum is somehow not as real as a normal distance can be refuted. If for example Achilles gives not 500 metre head start rather 900, then the reach will happen at $0.99999\dots$ which is the same as 1 that is a kilometre.

An even more effective infinite sum is using the repeated halves of the unit $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$

Then the leftover to the full 1, up to a member is exactly the last member that is:

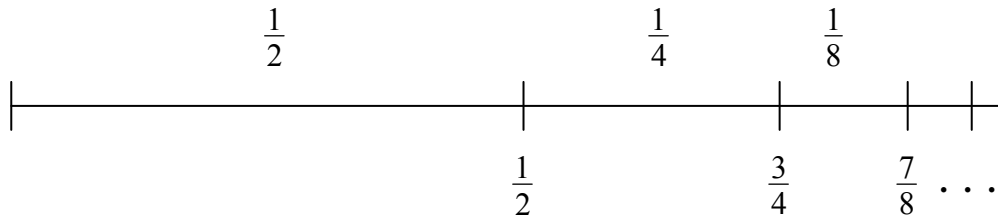
$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^n} = 1 - \frac{1}{2^n} \quad \text{Indeed,}$$

$$\frac{1}{2} = 1 - \frac{1}{2}, \quad \frac{1}{4} = \frac{1}{2} - \frac{1}{4}, \quad \frac{1}{8} = \frac{1}{4} - \frac{1}{8}, \dots \quad \text{Using these for each member we get :}$$

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^n} = 1 - \cancel{\frac{1}{2}} + \cancel{\frac{1}{2}} - \cancel{\frac{1}{4}} + \cancel{\frac{1}{4}} - \cancel{\frac{1}{8}} + \dots + \cancel{\frac{1}{2^{n-1}}} - \frac{1}{2^n}$$

This of course implies that the leftover diminishes to zero, that is : $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$.

All this can be seen visually too:



When these finite infinities become “natural” then subconsciously we fall into the trap of believing that all this was merely a simple consequence of diminishing numbers always adding up to finite. But this is false! So we have a new completely opposite Anti Achilles paradox too: Diminishing numbers can add up to infinity.

To show this lets use again the halvings, but with more and more repetitions, namely doubling their appearance too:

$$\frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}_{\frac{1}{2}} + \underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{\frac{1}{2}} + \underbrace{\frac{1}{16} + \dots + \frac{1}{16}}_{\frac{1}{2}} + \frac{1}{32} + \dots = \frac{1}{2} + \frac{1}{2} + \dots = \infty$$

We could say that we cheated because we didn't really use diminishing numbers due to the repetitions. So now we replace the repetitions with all different and diminishing values. Indeed, replace the first $\frac{1}{4}$ with $\frac{1}{3}$ which is even bigger. Replace the first $\frac{1}{8}$ with $\frac{1}{5}$, the next with $\frac{1}{6}$, then $\frac{1}{7}$ but leave the fourth $\frac{1}{8}$. Similarly, it will continue with $\frac{1}{9}, \frac{1}{10}, \dots$

So this increasing of the sum will use exactly all the reciprocals and be infinite in total too:

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \dots = \infty$$

7. Without the decimals: The problems with fractions

Avoiding the decimals, that is sticking to fractions is a very good practice in elementary school, that sometimes is missed due to the easy use of calculators. The fundamental problem with fractions is that they can be expanded, but that still keeps their value, for example:

$\frac{2}{3} = \frac{4}{6} = \frac{6}{9} = \dots$ So expansion means multiplying the top and the bottom with same number. But, equality doesn't mean that one is expansion of the other, because they might be expansion of a third. For example above, $\frac{4}{6} = \frac{6}{9}$ and they are not expansions of each other.

This fundamental problem has a formal meaning, namely: How can different things be equal?

We already had a similar phenomenon by converting numbers into other base: $2008 = 5566_7$

But here, with fractions, there is no such added base symbol that would make the equality of different meaningful. So some pupils already will have a subconscious aversion to fractions.

The other more practical problem is simplification. While, it is very easy to expand a fraction with a certain multiplier, it is hard to see if a fraction can be simplified or not. We are not aware of this because we only use small numbers in fractions.

In some schools they mention that a universal method of simplification is by prime factorizing

the numerator and denominator. Indeed, if $\frac{a}{b} = \frac{p_1 p_2 p_3 \dots}{q_1 q_2 q_3 \dots}$ then all we have to do is cross out

the common prime factors from the top and the bottom.

Example: $\frac{60}{100} = \frac{2 \cdot 2 \cdot 3 \cdot 5}{2 \cdot 2 \cdot 5 \cdot 5} = \frac{3}{5}$

Unfortunately, in absolute no school do they mention that this method is hiding a double or triple mystery. Indeed, what if a and b would have a d common divider, but it wouldn't show up in the prime factors. Of course, the hidden assumption is that the prime factorization of d must be part of a 's and similarly part of b 's. But it could very well be that d 's prime factorization is not appearing completely in a or b . Or it also could be that we are very lucky because d is product of primes in a 's and b 's factorization but these primes are not all the same as d 's own factorization. So a seemingly secondary hidden assumption is that a number can't be the product of primes in different ways. Then the prime factorization we introduced going through the minimal dividers is the only prime product giving our number.

Luckily this uniqueness of the prime products proves at once the main assumption in the simplification method, that is that a d divider of a has its prime factors all among a 's.

Indeed, let $a = p_1 p_2 \dots$ and $d = r_1 r_2 \dots$ and $\frac{a}{d} = \frac{p_1 p_2 \dots}{r_1 r_2 \dots} = c$.

Suppose $r_1 r_2 \dots$ would not all appear among $p_1 p_2 \dots$ that is, simplifying with the ones

that do appear, we would have $\frac{a}{d} = \frac{q_1 q_2 \dots}{s_1 s_2 \dots} = c = t_1 t_2 \dots$

Then, $q_1 q_2 \dots = s_1 s_2 \dots t_1 t_2 \dots$ would be alternative prime products because $s_1 s_2 \dots$ don't appear among $q_1 q_2 \dots$

But that's not all and that's why I mentioned above a triple mystery. Indeed, a third hidden assumption is that fractions that cannot be simplified and so in short could be called simple, are all unique. In other words, two simple fractions cannot be equal. Of course, we already mentioned that equal fractions don't have to be simplifications, like above $\frac{4}{6} = \frac{6}{9}$. But, this

new claim means that simplifying the two sides, they become identical. In other words, equal fractions are always expansions of a same simple fraction.

8. Indirectness: Walking in impossibilities

The two impossibilities above were, alternative prime products: $p_1 p_2 \dots = q_1 q_2 \dots$

and alternative simple fractions: $\frac{a}{b} = \frac{A}{B}$ This second at once implies the first. Indeed:

$a B = A b$ with prime factorization means: $p_1 p_2 \dots Q_1 Q_2 \dots = P_1 P_2 \dots q_1 q_2 \dots$

Here, the p -s and q -s can't have common member, since $\frac{a}{b}$ was simple, that is a, b had no

common divider, except 1. Similarly, $\frac{A}{B}$, so the P -s and Q -s are distinct too. But then, the

two sides could only be identical if all the p -s are exactly the P -s and q -s are the Q -s.

Then of course, $a = A$, $b = B$ so we don't have alternative fractions at all.

The reverse, that is getting alternative fractions from any alternative $p_1 p_2 \dots = q_1 q_2 \dots$

prime products is much harder. In fact, we only show the existence. The trick is to assume that $p_1 p_2 \dots = q_1 q_2 \dots$ is the minimal valued among all the hypothetical alternatives.

Then first of all, $p_1 p_2 \dots$ and $q_1 q_2 \dots$ are not merely alternatives, that is not exactly the same, but must be totally distinct. Indeed, if a p_i were same as a q_j , then we could divide both sides with this, and obtain alternative prime products with smaller value, contradicting the assumption that $p_1 p_2 \dots = q_1 q_2 \dots$ was the minimal valued alternative products.

But, actually even more is true. Not only, are the two sides distinct, but none of the members of one side can divide a sub product of the other side. For example, p_1 cannot divide $q_2 q_3 \dots$ and q_1 cannot divide $p_2 p_3 \dots$

Indeed, if for example, p_1 would divide $q_2 q_3 \dots$ then, $p_2 p_3 \dots = q_1 \frac{q_2 q_3 \dots}{p_1}$ and this,

$\frac{q_2 q_3 \dots}{p_1}$ had a new prime factorization as $r_1 r_2 \dots$. But, $p_2 p_3 \dots = q_1 r_1 r_2 \dots$ would

have to be alternative prime products, because q_1 is not appearing on the left.

Of course, this were again a smaller alternatives than the original $p_1 p_2 \dots = q_1 q_2 \dots$

With this extra feature of non sub-dividability then, from $p_1 p_2 \dots = q_1 q_2 \dots$ we get

$\frac{p_1}{q_2 q_3 \dots} = \frac{q_1}{p_2 p_3 \dots}$ as alternative simple fractions. Indeed, p_1 doesn't divide $q_2 q_3 \dots$

and q_1 doesn't divide $p_2 p_3 \dots$, which also means that these fractions cannot be simplified.

And of course, $p_1 \neq q_1$, so the two simple fractions are different.

This obtaining impossibilities from each other is a fundamental feature of mathematics.

Not only is our logic indirect, that is proving facts from refuting their impossibilities but it is playing with these impossibilities. Formalism can't explain this mental process.

What are the fragmented non existent realities we imagine, like alternative simple fractions and prime products? A more practical question is where does it all end? Obviously we exchange an impossibility to one that is easier to refute. Here, the more difficult to obtain from the other was the alternative simple fractions. This already suggests that this would be easier to refute.

A more important consequence of walking in impossibilities can be realizing that the impossibilities are too beautiful to be rejected and instead we should look at wider realities. This is how non Euclidian geometry was discovered. Janos Bolyai wrote: "From nothing I created a new different word".

These indirect arguments and examined impossibilities are definitely not the didactical approaches into math. Luckily there is always a golden direct road. In fact, even the seemingly smooth but hidden traps can be eliminated.

The fractions are exactly such trap. So even though above I complained about elementary schools not mentioning the hidden mysteries behind the simplifications, this doesn't mean that I advocate solving those mysteries through fractions. They simply should be mentioned but not resolved. Later in high school they can be attacked but starting from the bottom. Then it turns out that though the impossibility of alternative simple fractions is much easier to prove than the impossibility of alternative prime products, still the primes should be the start not the fractions.

9. A questionable analogy: The atomness of primes **A heuristic generalization: Relative Primes**

From the definition of primes as non composites the word atomness is quite logical. The Greeks used the word atom for hypothetical undividable units of matter. Later the word was reintroduced by chemists and even later it turned out that the elements or different atoms are actually made up from smaller particles, protons neutrons and electrons. In fact the protons and neutrons are themselves made up from even smaller subatomic particles. More importantly, the atoms can combine into molecules and molecules into more complex molecules again. If we ignore the inside of the atoms then they are an analogy to the primes. Namely the molecules can be regarded as the products and the atoms as prime factors. Just as the atoms remain themselves in the molecules, the primes remain units in products. First of all the claim is not quite true about the atoms again because they will share electron clouds, in fact that's the whole essence of chemistry. Anyway, the central parts the nuclei remain unchanged. But for the analogy to make sense, we have to tell in what sense will the primes remain themselves. If we cut a molecule into two then a particular atom must go either way, into the first or the second. If instead of a particular individual atom we merely regard a kind of atom then knowing that it was present in a molecule, the splitting will mean that at least one half must have that kind, but now both can have it too. Numbers are this "kind of atoms" not particular ones because many numbers can contain a same prime as factor. So now the claim is clear: If a p prime divides a $k = m n$ product, then p must divide m or n maybe both. The primes are simply the numbers that are not composite, that is have no factors. So they are internal atoms and so our claim simply means that this implies external atomness too. The reverse is quite obvious, that is external atomness implies internal at once. A number that divides all products separately can not have factors itself. To put it negatively, if a number is a composite $c = a b$ then there are products $m n$ so that c does not divide either m or n and yet c divides their product. Indeed, c already divides itself but neither a or b because they are smaller. But we can make also products that are bigger and so c will be a factor of it. We simply have to increase one in $a b$ say a , by multiplying it with any q so that $q a$ is still not dividable by c . Then $m = q a$ and $n = b$ gives such product. None of them is dividable by c but $m n = q a b = q c$ is. For example $6 = 2 \cdot 3$ and we can multiply 2 with many numbers like $2, 4, 5, 7, \dots$ so that it won't become dividable by 6 and thus the new products $4 \cdot 3, 8 \cdot 3, 10 \cdot 3, \dots$ will be all dividable by 6 but none of their two factors are.

Observe that internal atomness that is primeness itself is a finite affair. We only have to check the numbers less than p to verify it. But external atomness that is dividing products separately is infinite many claims. This explains why it is much harder to prove that internal implies external. In spite of these vivid atomic pictures, for the actual proof we should go back to basics and avoid the concept of dividability at all. What does it really mean that p divides the $m n$ product? Well, simply that $c p = m n$ so there is an equal product that involves p .

Beside this step back to multiplication, the other crucial step is a generalization from the p prime to any d divider of $m n$ that is to the $c d = m n$ product equation.

Many times it's easier to prove something more general. But here, it's not the case!

The separate dividability is only true for primes as dividers. So then how can we generalize? Well, we change the conditions too! Namely, from arbitrary $m n$ products, we go to arbitrary in only one member, say m , while the other, n is fix and required to have some new conditions with d . Since m is the arbitrary, the new more specific claim instead of d dividing one of the members will be that d divides m .

But what should be the crucial condition about n and d ?

It is that they have no common factor.

An other name for this is that they are relative primes.

A p prime is not necessarily relative prime with everything, because p is a factor of itself.

So the multiples of p have p itself as common factor with p .

Now we see why it was logical to regard a number as its own factor if it is not 1.

On the other hand, 1 is relative prime with anything because 1 has no factor at all.

In spite of these facts being in conflict with our feelings about primes, the truth is that relative primeness is a more basic concept than primeness.

For $d = p$ primes, this new rule at once gives the external atomness or separate dividability.

Indeed, p has no other factors at all, so being relative prime with n means that p doesn't divide n . This implies that p divides m and so separate dividability stands.

Observe that this new rule is still going from finite to infinite. Indeed, for any d and n , we can check all smaller numbers to see that they have no common factors, that is are relative primes.

10. Minimality: Proving the atomness of primes through Relative Primes

The crucial trick of our proof is to forget even about our new conditions of d and n .

Just fix arbitrarily them and then regard all possible c, m pairs that make the $cd = mn$ equation possible. One trivial solution is $c = n$ and $m = d$ because: $nd = dn$

But actually all k multiples of this are solutions too: $knd = kdn$

If c, m are a solution and we increase or decrease c to a c' then the left side cd obviously increases or decreases too. So, to keep equality m has to increase or decrease too.

Thus if $cd = mn$ and $c'd = m'n$ are two solutions, then the relation of c, c' and m, m' are the same, so if $c < c'$ then $m < m'$ too and vice versa. This means that among all the possible c, m solutions, the order is the same, either by the c -s or the m -s.

But these are natural numbers, so there has to be a minimal c_0 , among the c .

The corresponding m_0 is also minimal among m because of the common order.

Obviously, again $2c_0, 2m_0$ and $3c_0, 3m_0$ and so on, are solutions too. We claim that:

There are no other solutions! That is, all solutions are kc_0, km_0 with $k = 1, 2, 3, \dots$

Since the order of c, m are the same, if there were a possible c between

kc_0 and $(k+1)c_0$, then the corresponding m were also between km_0 and $(k+1)m_0$.

Of course, under the minimal $k = 1$ case there can be no solution.

So any other solution apart from the above sequence can only be:

$c = kc_0 + \gamma, m = km_0 + \mu$ with $\gamma < c_0$ and $\mu < m_0$.

But this means that γ, μ is not a solution, so $\gamma d \neq \mu n$.

On the other hand, kc_0, km_0 are solutions, so $kc_0 d = km_0 n$.

Adding this to the non equality, the sums must remain non equal, that is:

$kc_0 d + \gamma d \neq km_0 n + \mu n$ or $(kc_0 + \gamma) d \neq (km_0 + \mu) n$

So $kc_0 + \gamma, km_0 + \mu$ are not solutions either.

The fact that all solutions are kc_0, km_0 means that apart from $k = 1$, that is the minimal c_0, m_0 , all other solutions have a k common factor.

On the other hand, the minimal c_0, m_0 solution always must be without common factor.

Otherwise we could divide both, getting a smaller solution.

So the minimal solution is the single relative prime solution.

If d, n are relative primes, then the $c = n, m = d$ trivial solution is a relative prime one, and thus it must be the minimal one as well.

So, all solutions are $c = kn, m = kd$ and so d must divide m .

We went from p primes to d dividers that are relative primes with n .

We keep this assumption but now in a true sense, generalize our claims.

11. Remainders: A second proof with more details

The claim that if d divides $m \cdot n$ then it has to divide m , could be also said as:

If d doesn't divide m , then d can't divide $m \cdot n$. This should stand for any m , so we let m go through all numbers, and claim that all $m \cdot n$ products, where m is not multiple of d will not be multiples of d either. This is a step towards infinity, but where is the more general claim?

Well, $m \cdot n$ not being a d multiple, that is d not dividing $m \cdot n$ also means that $m \cdot n$ when divided by d , should have a positive remainder. The remainder of a number to d will be abbreviated as $[\]$ and $m \cdot n$ not being dividable by d , simply means that the remainder is not 0, that is positive: $[m \cdot n] \neq 0$ rather $[m \cdot n] = 1$ or 2 or \dots or $d-1$.

So with all conditions, our old claim is simply this:

If d, n are relative primes, then $[m \cdot n] = 1$ or $2 \dots$ or $d-1$.

With our fix n we'll step through all m values and will denote $[m \cdot n]$ as r_m .

$n, 2n, \dots, (d-1)n$ are the first $d-1$ many $m \cdot n$, having the remainders, r_1, r_2, \dots, r_{d-1} .

The next $m \cdot n = d \cdot n$ of course has 0 remainder, so $r_d = 0$.

Then $(d+1)n, (d+2)n, \dots, (2d-1)n$, will have again the same remainders, that we had from the start because $(d+k)n = d \cdot n + k \cdot n$ and the an addition of $d \cdot n$ doesn't change the remainder. So with indices $r_{d+1} = r_1, r_{d+2} = r_2, \dots, r_{2d-1} = r_{d-1}$.

Then again comes a definite 0 remainder of $2d \cdot n$ as r_{2d} .

And so on, we keep repeating the same initial block. In fact, we may find this indexing of all remainders up to infinity insane. But we'll see soon how useful this is.

Right now, we have to make our more general claim. The old claim simply is that the repeating block of remainders can not contain a 0 value, except at the end. The new, much stronger claim is that the block r_1, r_2, \dots, r_{d-1} contains all possible values, except 0.

So they are merely a re-ordering of the $1, 2, 3, \dots, d-1$ numbers.

This re-ordering is quite unpredictable from d itself, so there is no formula that would give it for all possible d values. But most amazingly, the fact that all $1, 2, \dots, d-1$ values appear, follows easily from the infinite continuation.

In fact, three simple rules will be enough to prove our claim.

But first we have to see what the r_m remainders really are. Well, $r_m = [m \cdot n] = m \cdot n - c \cdot d$, where c is how many times we can multiply d without going above $m \cdot n$. If $c \cdot d = m \cdot n$ then r_m is 0, if $c \cdot d < m \cdot n$ then r_m is positive but $< d$, because we went as far as possible up to $m \cdot n$. If r_m is small, we can double or triple m and c and thus, double or triple r_m itself.

Thus, r_{2m}, r_{3m} become $2 \cdot r_m, 3 \cdot r_m$. But this simple multiplication of remainders must stop, because no matter how small r_m is, a $k \cdot r_m$ outgrows the d divider.

For example, 7 has 2 remainder to 5. The double, 14 has double remainder 4, but already the third multiple, 21 has only 1 not $3 \cdot 2 = 6$. This virtual overgrown remainder 6 is still useful though, because it has the same 1 remainder. But there is an even more interesting use of the continued multiple virtual remainders. Namely, instead of looking where they reach d , the divider, we can also regard when they go over the n fix step we use. For example, above this is 7 and $3 \cdot 2$ was still under. But then, $4 \cdot 2 = 8$, goes over 7, having 1 remainder again. This is not a coincidence! The $4 \cdot 2 = 8$ virtual remainder is from $4 \cdot 7 - 4 \cdot 5$. Subtracting the 7 means exactly $3 \cdot 7 - 4 \cdot 5 = 1$, that is the third remainder being 1.

The three rules will be these three simple facts:

Multiplication of remainders if they are still under d .

Overgrowing d and thus, having a new remainder to d .

Overgrowing n and thus, having a remainder to n that is a previous n multiple's remainder to d automatically.

So now the precise rules with precise proofs:

1. If $k r_m < d$ then there is $r_M = k r_m$. Namely, $M = k m$ will do, because:
 $r_{k m} = [k m n] = [k [m n]] = [k r_m] = k r_m$.
2. If r_m doesn't divide d , then there is $r_M > 0$, that $r_M < r_m$.
 Namely, let k be the first number that $k r_m > d$ and then, $M = k m$ will do.
 Indeed, $r_{k m} = [k m n] = [k r_m] = k r_m - d > 0$ by $k r_m > d$.
 And since $(k - 1) r_m < d$ thus, $k r_m - d = r_{k m} < r_m$.
3. If r_m doesn't divide n , then there is $r_M > 0$, that $r_M < r_m$.
 Namely, let k be the first number that $k r_m > n$ and then, $M = k m - 1$ will do.
 Indeed, $r_{k m - 1} = [(k m - 1) n] = [k m n - n] = [k r_m - n] = k r_m - n > 0$ by $k r_m > n$.
 And since $(k - 1) r_m < n$ thus, $k r_m - n = r_{k m - 1} < r_m$.

The 2, 3 rules can be combined by saying:

If r_m doesn't divide both d and n then there is $r_M > 0$, that $r_M < r_m$.

With our conditions, that the d, n are not multiples of each other and have no common factors either, the only number that divides both of them is 1.

Thus, for any $r_m \neq 1$ there is $r_M > 0$, that $r_M < r_m$.

This means that eventually there has to be an $r_M = 1$.

The first rule then says that all multiples of this 1, that is all numbers up to d , are remainders.

12. Big pictures, big circles: Linear combinations, Unique Prime Factorization

The appearance of all possible remainders for the $m n$ multiples was interesting but didn't give any useful extra fact about primes or factors. It merely gave new proof for the external atomness of primes if we regard the d divider as a prime. Amazingly, if we generalize further by not restricting the n starter number and d divider at all, that is regarding the remainders for any $m n$ multiples, then we can learn some new and practical facts.

The 1, 2, 3 rules remain valid but now we seemingly can not use them because we don't know that the only common divider of d and n is 1. But we can start with something obvious instead. Namely that among the remainders still has to be a smallest r . Then rule 1 tells at once that all the remainders are merely the $k r$ multiples of this minimal. The combined rule 2 and 3 will also tell that r must be a common divider of d and n . Indeed if it weren't then there had to be a smaller divider. But that's all! So we can still wonder which of the common dividers is this minimal r . Can for example other common dividers be among the other $k r$ remainders? Something obvious is lying in front of our eyes and yet we simply ignore it. Most importantly this obviousness is a door to a whole new vision. What is a remainder in general?

It is the leftover from subtracting the divider. So a remainder of d in n is $r = n - c d$.

We used this already above but not for the n itself rather for its multiples that is as:

$r_m = [m n] = m n - c d$. This is actually a hint towards the bigger picture. Indeed, using the more general multiples of n , instead of the fix n itself, made the difference more symmetrical by allowing the m and c multipliers for both n and d . We could reverse their roles too and this was a hidden feature in rule 3 already. In fact, the mentioned virtual remainders simply meant that we don't pursue the restriction of using the highest possible c value that still keeps the difference becoming negative that is $c d$ overgrowing $m n$. The next step would be to allow even negative differences and finally to allow not only differences but sums as well:

$q = m n \pm c d$. This completely departs from the original meaning of remainders that's why I used the new q letter. Indeed, using sums will create huge numbers, so seemingly we go against the very purpose of remainders. But the remainders are still part of these q numbers.

The official name of these q are "linear combinations" from n and d . The word combination is obvious and linear simply means that we only used multiplication.

Now comes the obvious fact that I mentioned:

Every common divider of n and d remains divider of any linear combination from them.

Indeed, multiples, sums or differences always keep the common dividers. But then since all the remainders of d in $m n$ are also linear combinations of n and d , thus these r_m all must be dividable by all the common dividers of n and d too. This includes the minimal r too, so r must also be dividable by all common dividers of n and d . This at once shows that the other $k r$ remainders can not be common dividers because r is not dividable by $k r$. But more importantly r is a common divider of n and d , so it is actually a super common divider of them, meaning that all other common dividers divide it. Being “super” of course at once means being the greatest but it’s even more because we could imagine the greatest among the dividers without having all others as dividers. But our result is now totally general for any n and d . So in fact we just proved that the r minimal remainder is always creating this super common divider. So in fact all greatest common dividers are actually super.

We can feel that this superness of all greatest common dividers is a grand result.

Seemingly even quite independent from the infinite sequence of alternative products and from its refinement the infinite sequence of r_m remainders. But clearly it must relate to these backwards too, in fact also to the original problems of alternative that is equal fractions, to the atomness of primes and to the impossibility of alternative prime products.

So now we connect these and thus see that they are all the same.

A common divider of n and d is at once a simplifier of the $\frac{n}{d}$ fraction.

The greatest common divider is the greatest simplifier, so it gives the “smallest” simplification.

This of course is not smallest in value which is the same as $\frac{n}{d}$, rather smallest in numerator and denominator. So this “smallest” is best called minimal.

The greatest simplifier being super, means that all other simplifiers divide it and so all simplifications are expansions of the minimal simplification.

This easily proves that equal fractions are always multiples of each other or a third, as follows:

For any two $\frac{n}{d} = \frac{c}{m}$ equal fractions we can find a common, that is identical expansions of

them, namely as $\frac{n c m}{d c m}$. Indeed, this is an expansion of $\frac{n}{d}$ but also of $\frac{c}{m}$ because $n m = c d$.

But then for this common expansion we can apply our result that all simplifications of it are expansions of the minimal. This includes the two equal fractions which are then indeed expansions of or itself the minimal.

Observe that here we used minimality again but only among the simplifications of a fix fraction. So we avoided a direct use of infinities and also indirect assumptions of non multiple variants.

Now we show that the already twice proved atomness of primes also proves the superness of the greatest common divider. We simply have to close the circle by going back to one of our original mysteries, the impossibility of alternate prime products $p_1 p_2 \dots = q_1 q_2 \dots$

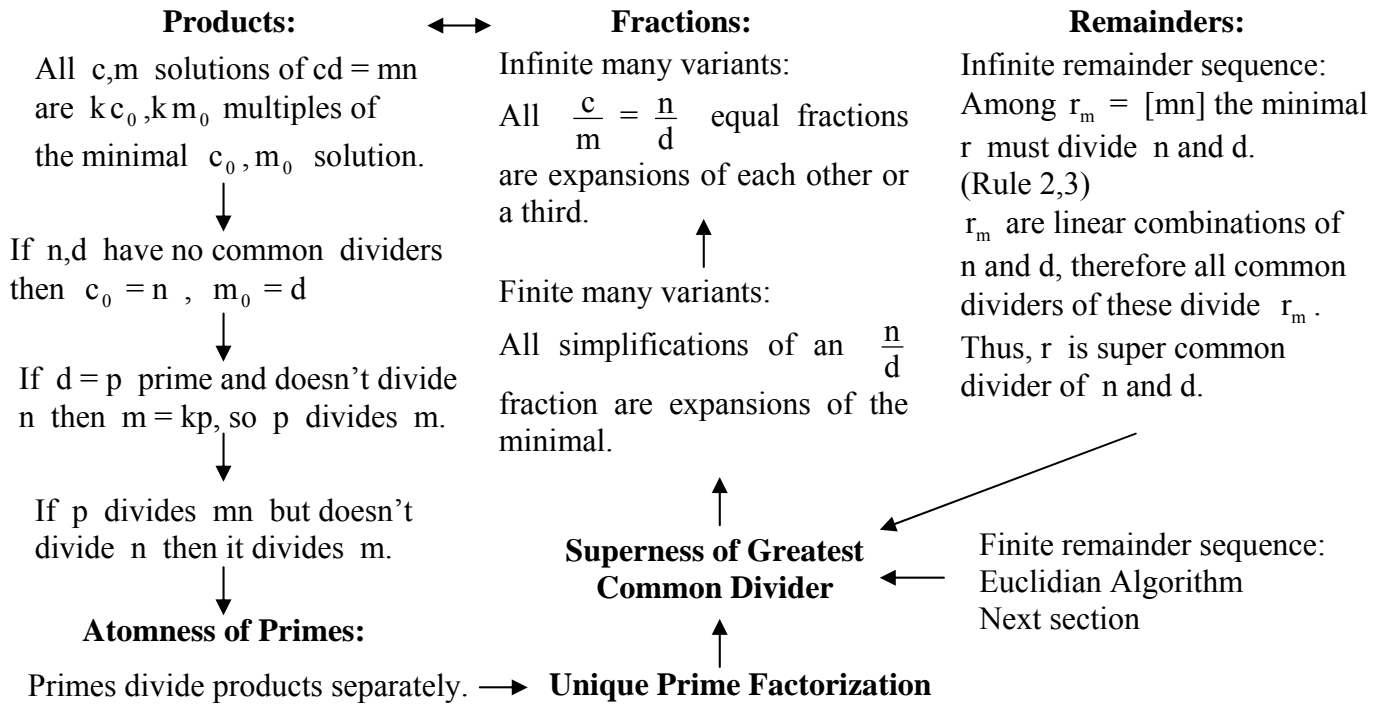
Or as it is called in a positive manner: the Unique Prime Factorization.

The proof of this by the atomness of primes is easy again! Every prime on one side must divide a factor on the other side, but since they are primes too, this means that it must be one of them.

Then we can divide both sides with this and repeat the argument again and again. Finally equal primes must remain, so both sides were already the same at the start. This also means that the original increasing prime factorization that we introduced, is the single one.

The Unique Prime Factorization also means that any factor of a number can only have prime factors from the prime factors of the number. In fact, all factors are merely picking some prime factors from the total. So then a common factor of two numbers is simply picking common prime factors with common appearances. The greatest common factor is simply picking all common prime factors with highest common appearances. So the common factors obviously must divide this greatest, and thus the greatest is automatically a super.

Table of Logic among Concepts



13. Theory and practice, The Euclidian Algorithm

The linear combinations were a great theoretical step from remainders but the root of our arguments still relied on the minimal r remainder. The fact that this r is actually the super common divider even allowed a more concrete that is less infinite and indirect line of proofs. But we want more and try to avoid the r as defined by the infinite remainder sequence. Indeed, this r is the super common divider of n and d so there might be a formula that gives r .

This would be very practical too because the r is also the super simplifier of the $\frac{n}{d}$ fraction.

Well, there is no formula, only a fast direct method to obtain r from n and d .

This method proves again that the best practice originates from the right theory!

But of course such right theory must be more than empty abstraction, it has to be new vision.

And indeed, even though the linear combinations were the right vision to see that the common factors inherit to them, we still missed an other "forest from the woods". There is something else just as important and simple about the linear combinations and it will have the totally practical application. What is this simple fact we missed? It is that:

Linear combination of linear combinations is again linear combination!

Indeed, from $n = 7$ and $d = 5$ lets make two combinations: $2 \bullet 7 + 1 \bullet 5 = 19$ and $4 \bullet 5 - 2 \bullet 7 = 6$ Then from these two a new one: $1 \bullet 19 - 3 \bullet 6 = 1$

Simply writing the "defining" combinations into the values, we at once get a direct combination from 7 and 5 obtaining the same 1 value: $1 \bullet 19 - 3 \bullet 6 =$

$$1 \bullet (2 \bullet 7 + 1 \bullet 5) - 3 \bullet (4 \bullet 5 - 2 \bullet 7) = 2 \bullet 7 + 1 \bullet 5 - (3 \bullet 4 \bullet 5 - 3 \bullet 2 \bullet 7) =$$

$$2 \bullet 7 + 1 \bullet 5 - 12 \bullet 5 + 6 \bullet 7 = 8 \bullet 7 - 11 \bullet 5 = 1$$

There are simpler ways to get the 1 value too, as $3 \bullet 5 - 2 \bullet 7$ or $3 \bullet 7 - 4 \bullet 5$, but the point was merely that combination of combinations is a combination. So this indeed was a purely theoretical proof but now we return to remainders.

What makes the remainders special among the linear combinations is that they are always definitely smaller then both the divided n and the dividing d . In our original application this advantage was sidelined because we kept on multiplying the n with increasing m values.

Our three rules did give an effective method to find the minimal r but only through using huge $m = M$ multipliers that is indices. Now the big new idea is this:

Lets keep calculating always the remainders from the last two combinations.

So for example starting again with 7 and 5 the first remainder is $7 - 5 = 2$

Next we use this with 5 and so the next remainder is already $5 - 2 \cdot 2 = 1$

This is obviously the smallest possible non zero remainder and so indeed 7 and 5 have no other common dividers. A more interesting example is 72, 42

The remainder is $72 - 42 = 30$ and the new pair is 42, 30

The remainder is $42 - 30 = 12$ so the new pair is 30, 12

The remainder is $30 - 2 \cdot 12 = 6$ the new pair is 12, 6 and they have 0 remainder.

The last non zero remainder was 6 so this is the super common divider of 72 and 42.

The common factors of 72 and 42 are: 2, 3, 6.

The problem with such “super” practical methods like this to get the super divider is that they usually hide a lot of messy details. Since they work so well, they give us beliefs in things that are not logically following from the method directly. Here, I started with the grand vision of combinations leading to new combinations, so we might think that this is the only hidden part. Every new remainder is actually a linear combination of the original two numbers. This is important because this implies that the remainders must all have the common dividers of the original two numbers inside them as dividers. Also, the crucial innovative idea to calculate new remainders means that they have to be decreasing numbers, so they have to reach a minimal. But then the rest starts to be foggy. First of all, why does the minimal have to be zero? That’s still quite easy to answer. Indeed, a non zero remainder means that we can still continue the method, we can subtract it from the previous remainder as many times as possible. The method only stops at zero because subtracting that would give the same as the previous. But this brings in a new fogginess. The clear thing is that the remainders being linear combinations of the originals keep all common dividers, but will this stop at zero too? And which part, the being combination or the having the dividers? Well, the first obviously stays because zero is always an obtainable combination from any n and d trivially as $d \cdot n - n \cdot d$. On the other hand we might think that zero shouldn’t have any dividers. So the fogginess turned into a nightmare because then our grand claim about the combinations keeping the dividers is not even true. But there is no need to panic, philosopher is on board! The non zero linear combinations do keep the common dividers. Zero actually only enters as a number through the combinations and we are free to allow dividability of it if we want to. This is what formalism does too but without mentioning the passage from reality to abstraction or wider new reality. So the pope who claimed that the zero is the devil’s work may have been right to some degree! At any rate, without even allowing zero, the last remainder is still the same that we had as the last non zero remainder and claimed to be the super common divider. But why? It contains all common dividers, but why is itself a common divider? This doesn’t seem to follow from neither the grand vision of combination of combinations nor the particular use of remainders. Of course it must and we merely missed something. Lets start from the beginning! The first r remainder is $r = n - c \cdot d$ but this also means $n = c \cdot d + r$ so the initial n is a linear combination of the next two members in the decreasing numbers. This remains true for all members.

For example in our last example, these combinations of the members from the next two will be:

$$72 = 42 + 30, \quad 42 = 30 + 12, \quad 30 = 2 \cdot 12 + 6, \quad 12 = 2 \cdot 6 + 0$$

The last two remainders were 6 and 0 so these have no following two members, they are merely the members in the previous remainder 12. But 0 being there simply means that this previous to the last non zero remainder 12 is dividable by the last non zero 6.

Now we can apply the grand vision backwards, that is say that all the earlier members are linear combinations of the last two non zero ones. So they keep the common dividers of the last two.

But the last one divides the previous, so the last one is actually such common divider of the last two and thus must divide all earlier including the initial two numbers.

14. Fractions: A Field of Actions, A new Silver Platter and three algorithms

I have to start with an admission. These last few sections in the book have been rewritten because I went through a major change in my philosophy. In fact I called this the fifth big turn in my life and it initiated a whole new introduction to Cognum.org. This is not the place to talk about those crucial things, only how they affect the didactical correctness in math. In spite of this, I want to be precise about the matter so I have to state it on solid philosophical grounds.

My fundamental mistake before lied in a naïve separation of actions and thoughts.

Fist of all, actions must include potential actions too. The so many times mentioned silver platters for example are not merely visual explanations. They have to work and can only “work” because we work with them. Even the most passive viewing of an explanation means a subconscious playing with the visions. To pursue the actions is already an act of math.

Seeing that the division of two numbers gives a periodic decimal is not the same as doing such divisions. In a sense, all levels of thoughts correspond to levels of potentialities in actions. Namely, how realizable the actions are. The fundamental claim of yoga that all actions are lies is still true but to proceed to higher thoughts, we have to enter actions. We have to lie to get more truth! Mathematics is a particular field of actions in thoughts.

Before, I always emphasized that math is simply common sense pursued with uncommon precision. But this only reflects the visions and there is a corresponding field of actions. This is a hidden world to us yet. Only mathematical dreams reveal something about it. Actions are also connected to emotions and normally these control dreams. Mathematics is the rules of actions without emotions. This is a much harder definition to swallow.

But what do all these deep philosophical connection mean to learning math? Simply that understanding and consciousness can not be the only guiding line. In fact, elementary school is exactly the period where true understanding shouldn't be pursued and rather the common sense of actions should grow!

The first concrete world of these actions is the multiplication of numbers. The times table must be learnt by sheer practice and not by exploration with meanings. This fact was bitterly learnt by those experimenters who tried to train smart kids and ended up with innumeracy. But there is a singular actional common sense hiding behind the times table too, namely the exchangeability of multiplication order or professionally called commutativity.

The simple truth is that $3 \bullet 4 = 4 \bullet 3$ can only be seen in the plane as the tiles in a rectangle having sides 3 and 4. No combining of sets or proportionalities can avoid our limitations in time. The fact that the line or even just the natural equal distanced dots on it, provoke the two dimensional plane is the nature of things. All mathematical facts of a system beg for wider systems. And all mathematical understandings create wider mysteries.

At the same time as some morons in the west tried to eliminate the times table, in Russia a miracle was created by Larichev. His Collection of Algebraic Exercises not only teaches the arbitrarinesses of our earthly variable notation system but conquers the universal transition from reality to abstract by enabling anybody to write and solve equations for word problems.

I entered math through Geometry and only later when started tutoring did I realize that the word problems are the universal door to math. So elementary math education is a solved problem. It is merely a political problem today. Of course you can improve Larichev's book and explore other activities but the point is that the times table must be continued. Algebraic numeracy must be learnt by sheer practice of conquering reality.

It's not a general relaxed or not demanding attitude that stops the practice of word problems till a point of ease and fluency. Indeed, so much abstract nonsense is required already in elementary school. Crazy names and categories, verbal answers without having correct visions, multiple choices, percentage levels of “mastering” skills. The system is rotten to the core. Elementary school has become the beginning of all social lies. Kindergarten is turning into the same as we speak. At high school level understanding should be the guiding line. But it is crucial for an other reason too, namely this is where socialization freezes. So society is manipulating this level of education the most for its own goals. But we have to return to fractions which are obviously lower elementary school level.

I already mentioned a tendency to avoid fractions at all. This is damaging algebraically as avoiding exact solutions of equations. For word problems these exact solutions are even more important. I also mentioned above that fractions as actual division process influence the silver platter of decimals. All these are the fractions as the field of actions referred in the title.

But there is a geometrical importance too. This is the new silver platter referred in the title.

It is the crucial transition from numbers to the plane and to an effective exploration of the plane. To write anything on a blackboard is already an exploration of the plane so it can not be avoided. As I said, the only visualization of $3 \bullet 4 = 4 \bullet 3$ is by looking at a rectangle. By this time the pupil knows that both products are 12 so the “seeing” is merely a confirmation of an earlier blindly accepted wisdom. Division, that is fractions have their own wisdoms but here the consciousness is missing too and so the plane can be this door to realization in both sense as recognition and action. The same simple tiling of the plane where we can see rectangles with whole sides, is the background for the fractions. In fact, a blackboard with grids on it should be in every classroom. Most importantly this shouldn't be turned into a coordinate system!!!! No special horizontal or vertical lines as coordinates!!!! This tiled blackboard is a piece of an infinite monotonous grid system over the whole plane. Also, the natural numbers should never be represented as dots going from a fixed point of origin to one direction, usually to the right.

The coordinate conception is the worst anesthetic of minds at elementary school visualizations.

It's a disease spreading through generations contradicting the ancient pure drawings in the sand.

The natural numbers have many uses in an infinite gridded plane. They can be some many grid points, some long horizontal or vertical connectors, and even some many tiles. But all this must happen with the proper geometrical background because the naturals already contain the plane. Hidden, secretly waiting to be discovered. The shiftable and proportionally alterable pictures are the apriori certainty we all possess about space. Seeing is believing and the third dimension is merely action in abstractions. But the ancient pure drawings in the sand lacked a fundamental new extension, only recognized by Gauss. Just as Newton's grand vision of forces should be in elementary school, the grid connector multiplications of Gauss should be seen by kids too.

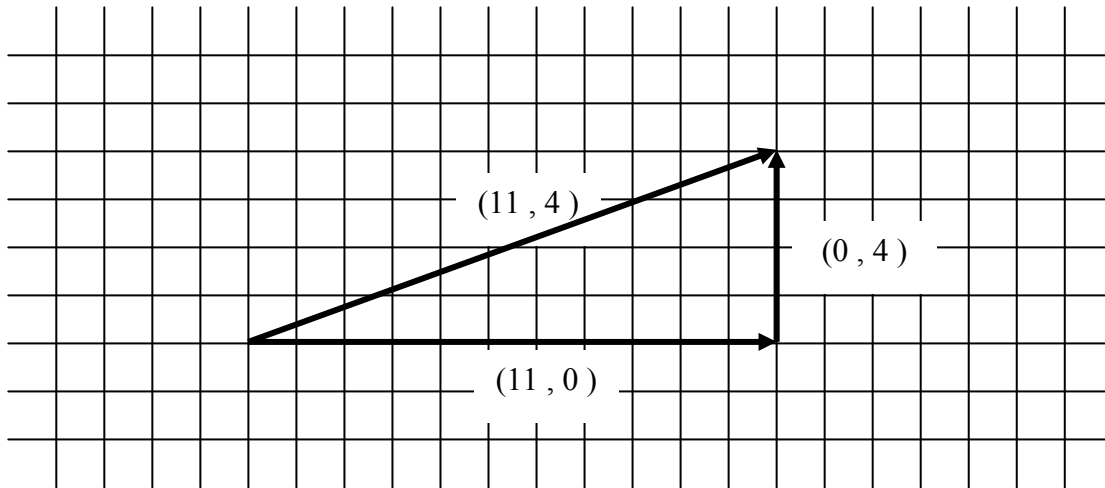
The fact that Newton and Gauss didn't initiate crusades to impregnate the education system and force, that is enable everybody to see what they saw, is a separate issue. Kant the philosopher did react to Newton properly and I spelled out my belief about the naturals containing the hidden wider realities. But this is still not enough. The political step is pinpointing the didactical skeleton. This is the only pure action that you not only can but must practice. If you are wrong in some details, it will turn out, but not even believing in an ideal transfer of knowledge is missing idealism at all. But there is worse than that. Namely believing that we shouldn't even think about this, pretending that there is no contradiction to start with. That knowledge is personal, that everybody is entitled to their beliefs, that there are small minds and great minds, that small minds simply should learn from great, that great minds should simply tolerate small. And the list goes on because you try to escape the simple truth that everybody can know everything. And everything that stops you from acting towards this, is simply a trap of your own mind. So I got carried away again but now really back to the plane!

An elementary introduction into the grids starts with the concept of “trip”. This means a first horizontal and then vertical move from any starting point. The trip itself is dictated by (x, y) , allowing any positive or negative numbers or zero for x and y . The agreement of positive meaning right horizontally and up vertically is introduced. Dictating different trips, the student can then execute the actual motions on the grids.

Next comes the “connector”. This is an arrow from a start grid to an end. Then we show that every connector defines a unique trip because the horizontal direction and step is determined and then the vertical again. We especially investigate the horizontal and vertical connectors that must involve zero in their trips. Finally we show that there was a trip that can not be shown by connectors namely $(0, 0)$. This should be denoted merely as an x crossing of any point.

Next we introduce two words simultaneously, the “coordinates” and the “components”.

The first is the values we used in the trip, the second is the horizontal and vertical connectors corresponding to them. So the coordinates are merely whole numbers, while the components are special connectors themselves:



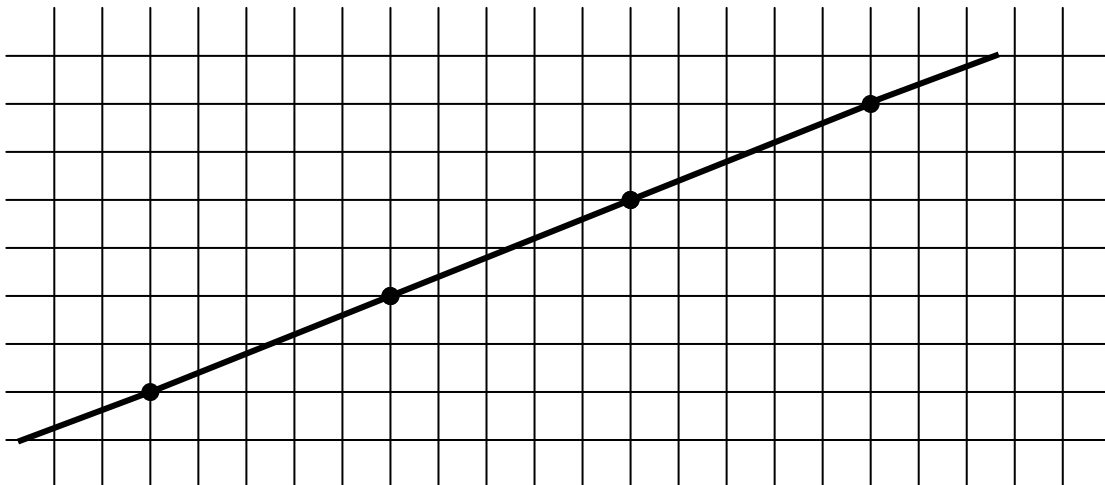
The next concept is the “multiple” of a connector. The name speaks for itself. We simply multiply both coordinates with a k common natural number, so form (kx, ky) .

This can be accomplished by sliding the connector in its line k many times. But both the original and multiple connector can be shifted in any direction too. So the simple universal fact is that the multiples are all parallel with the original and so parallel with each other too.

The reverse of the multiple is a simplification! So any (u, v) connector is a simplification of (x, y) if $(x, y) = (ku, kv)$. All this starts to resemble the multiples and simplifications of fractions. But don't even try to anticipate what comes next because it's a real miracle!

The next concept is the crucial in the miracle! A connector is “minimal” if there is no grid on it. Multiple connectors all have grids, so minimal is the exact opposite of being a multiple, that is having simplification. Among fractions, being simple that is not having simplifications is the natural concept and minimality was merely introduced for the very proof that the simple ones are unique. Here, this natural minimality among the connectors is forecasting the miracle.

The grids on an infinite line are simply the repetitions of the minimal connectors on it:



Now comes the miracle, an almost trivial chain of logic:

If two connectors (u, v) , (x, y) have a common multiple that is: $(ju, jv) = (kx, ky)$, then the two connectors are parallel and either one is simplification of the other or they have a common simplification.

The crucial element is the middle claim of parallelity that connects the common multiple to the common simplification. The parallelity itself is obvious because both connectors are parallel with the common multiple and thus with each other too. Then the two connectors can be shifted to a common line and thus the second part follows again. Indeed, the two will be both multiples of the minimal of the line. Except of course if one of them itself is minimal.

The application of the miracle is still a surprise: Having common multiples, for the coordinates means: $ju = kx$ and $jv = ky$. Dividing them we get: $\frac{u}{v} = \frac{x}{y}$ or $uy = xv$.

So we obtained the alternative fractions or alternative products from having common multiples.

But the real important fact is that in reverse too, the alternative fractions or products mean common multiples too! Indeed if $u y = x v$ then $(u x y, v x y)$ is not only an obvious multiple of (u, v) with $j = x y$ but also of (x, y) with $k = u y = x v$. So connectors with common multiples mean exactly the alternative fractions or products. And thus we obtained the result that they also have common simplifications.

The magic of the plane, namely parallelity, avoided the complicated proofs we used before.

Thus, parallelity also proved the atomness of primes and the Unique Prime Factorization.

But fractions have a more practical name for the connectors too, namely as their "slope". This is clearly relating by its every day meaning to the angle of the connector but it's a new concept.

The fact that it is merely an alternative for the angles is actually the whole problem that we attacked. But just as a start we should be at least roughly meaningful, so zero angle measured from the horizontal should be the zero slope.

This at once shows that the slope must be the fraction $\frac{y}{x}$ and not the opposite.

Indeed if y is 0 then we only go horizontally without a following up or down in the trip.

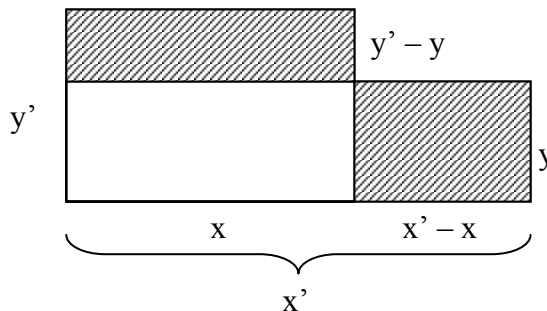
Also, a reversed, only "up" trip has no meaningful slope or should be infinity.

Same slopes of course mean equality of fractions and multiplying over giving equal products.

But this also means having common multiples and thus being parallel, that is having same angles. So indeed angles and slopes correspond.

As a weird exercise in concepts, we can go on without recognizing the heuristic concept of common multiples. Then the equality of products can be interpreted by equal areas of rectangles. But we can still go two ways. Merely recreate the old arguments with geometrical vision or bring in the plane for the areas. We'll show both:

1. Increasing or decreasing both sides of a rectangle, the area definitely increases or decreases. So to keep an area same, we have to increase one side and decrease the other.
2. If we alter the $x y'$ rectangle to a same area one, with increasing x to x' and decreasing y' to y then the $(x' - x) y$ rectangle is also same area with $x (y' - y)$.



For slopes, these two rules at once mean the followings:

1. If $\frac{y}{x} = \frac{y'}{x'}$ then both y' and x' must relate to y and x the same way.

That is, both are bigger or smaller than y, x .

2. If $\frac{y}{x} = \frac{y'}{x'}$ with the x' and y' being bigger, then $\frac{y}{x} = \frac{y' - y}{x' - x}$ too.

Now we can show that all equal sloped connectors are parallel!

Among the equal sloped connectors there has to be a minimal in its x coordinate as x_0 .

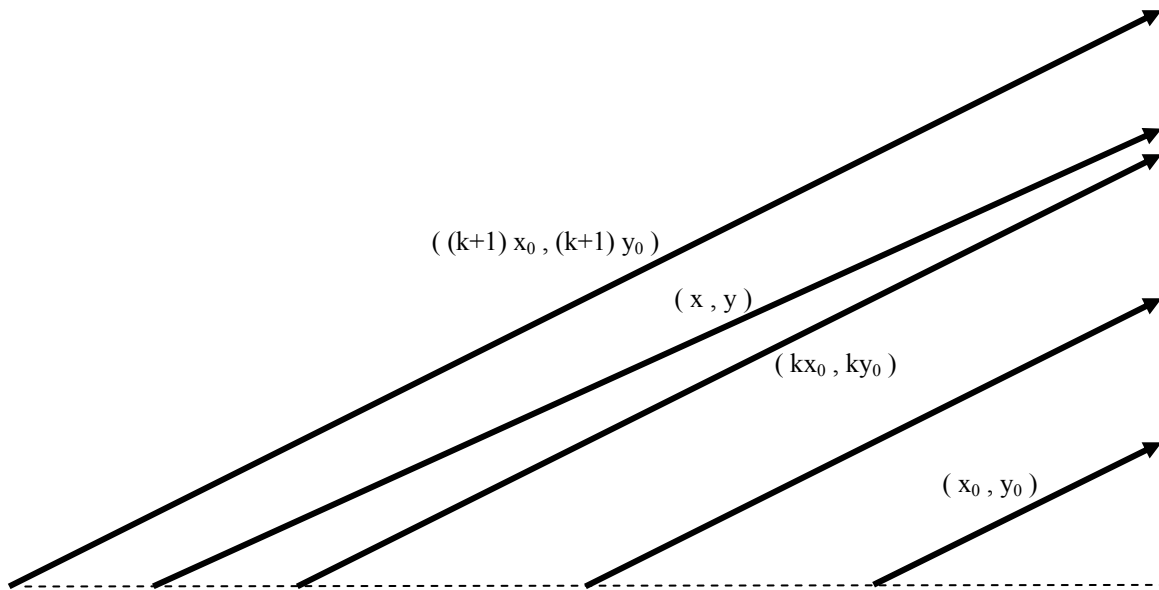
This connector of course will have the minimal y_0 too because equal slopes have same relations among x and y .

Now lets regard all $(k x_0, k y_0)$ multiples of this minimal (x_0, y_0) connector.

These are an infinite sequence of parallel connectors and we claim that these are the only ones having the same slopes. Suppose there were some other (x, y) .

The x_0 was the minimal so x can not be smaller than x_0 .

Any x value above x_0 and not being any $k x_0$ must be between some $k x_0$ and $(k + 1) x_0$ that is $x = k x_0 + u$ with $u < x_0$. The corresponding y then must be also between $k y_0$ and $(k + 1) y_0$ that is $y = k y_0 + v$ with $v < y_0$. So such impossible but now hypothetical (x, y) connector would have to lie literally between the k and $k + 1$ multiples:



The second rule about the slopes says that the difference between the $(k x_0, k y_0)$ and $(k x_0 + u, k y_0 + v)$ equal sloped connectors must also be same sloped.

But this difference is (u, v) and $u < x_0$ so it can not be same sloped.

Thus the assumption of an other (x, y) was false too.

We seemingly repeated the same arguments used earlier for the solutions of $c d = m n$.

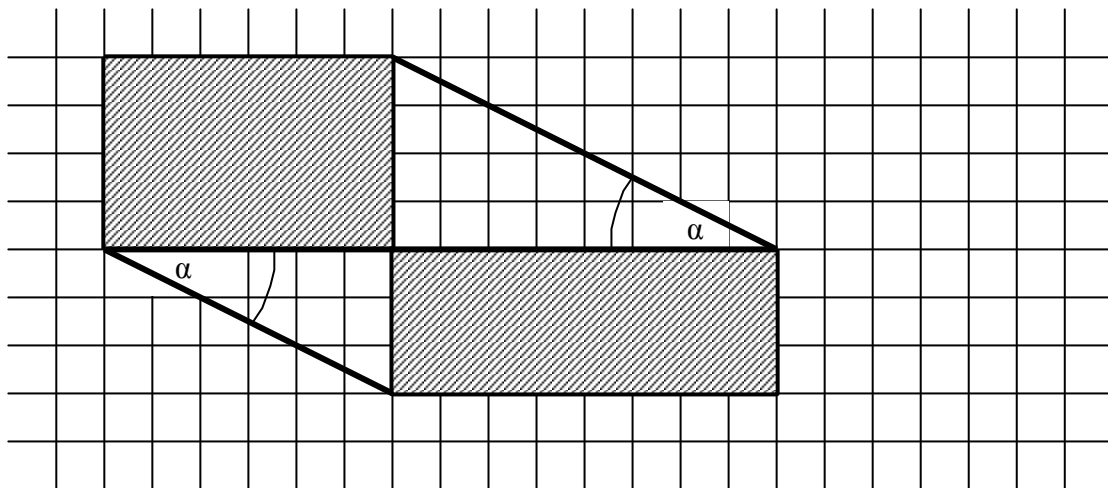
But now we had the visual meanings behind as the infinite many parallel connectors.

And yet the parallelity was only proved by excluding any “intruder” among the obviously parallel multiple connectors. If an intruder would be parallel then it were excluded too.

On the other hand, parallel connector lines are identical because shiftable over each other.

A better proof would be to use the rectangles to show that equal sloped connectors are parallel.

Not to interfere with each other, the best is to move the two rectangles touching in one common corner only. Then again the equality of the slopes means equal areas but the connectors and their angles are exactly visible too:

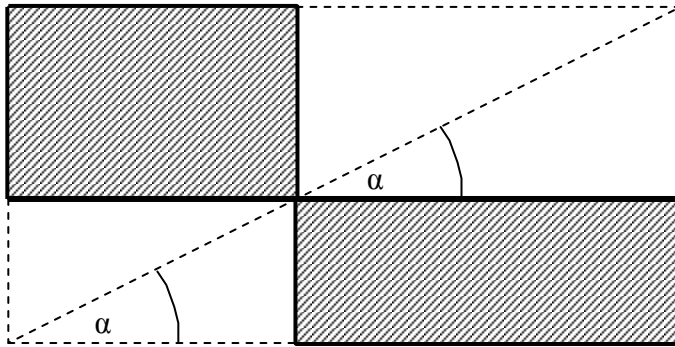


We prove much more! Namely first that the implication is reversible, that is not only equal areas imply equal angles but equal angles imply equal areas. And second that these are true among all rectangles of the plane not only grid ones.

As usual, we do this not to make the proof harder rather easier. So just regard the picture above without the grids in the background and this indeed will mean bringing in the background plane as an advantage.

First we show that if equal angles imply equal areas then non equal angles mean non equal areas, so the reversal of this direction follows. Indeed, if we fix three sides say the two horizontal and up vertical then the fourth down vertical is uniquely determined by measuring the second α angle down and thus cutting the vertical side line. The second area is then also determined and more importantly for different down vertical sides this area will be different because the other horizontal side is fix. So if equal angles create equal areas then non equal must create non equal areas.

Now to show that equal angles create equal areas, we merely have to complete the two rectangles into a big one. The two new small rectangles then will have an other diagonal than the one we used above for α . And of course these other diagonals will have the same α angles and thus will form a single line, that is a diagonal of the big rectangle:



Thus the equality of the rectangle areas is obvious! Indeed they are both the same large triangles minus two same small triangles.

These ridiculously complicated proofs show how simple our original connector approach was! In truth, connectors are the generalized natural numbers! This was discovered by Gauss but melted into his wider search towards the complex numbers. He felt a deeper physical meaning behind these and his instinct was correct. Today the complex numbers are the fundamental language of physics. Quite on the contrary, the didactical importance of his complex integers was never established. I was blind to it until few years ago and even then I made some serious errors in how to introduce them. Complex numbers are dragged in at tertiary level as some alternate reality but the Gaussian integers still remain outside like some useless toy. This is an insult to truth and understanding. These integers are exactly the same to the complex numbers as the naturals are to the reals. The only entrance, the unavoidable golden road. In fact they should be traveled before we go from naturals to reals. But to do this we shouldn't get lost in the delusional simplicity of coordinate systems! So we must use connectors! I will lead you onto this golden road exactly this way from section 23. That's where we introduce addition and multiplication of connectors. It should follow the concept of slope that didn't even need those.

But we are not going in an elementary school didactics now, so we have to pile up more conventional knowledge to make the connectors really show their power.

In this section we are focusing on fractions. They are deep, yet belong to elementary school. They are easy to use but hide a lot. Only in high school should we challenge the "why" completely but this doesn't mean we can not explore facts before that direct towards the "why". The heuristic common multiple argument for same sloped connectors, is definitely one that should be explored. This explores the plane, namely parallelity too.

Parallelity can be explored on its own! The three most trivial meanings of it are:

Having same distance between the two lines. = Fix distanceness.

Having same angle to a third line. = Same directionality.

Not crossing each other. = Non crossing.

The above introduced slope even gives a fourth definition as same slopedness.

But as we saw, it easily goes back to same direction.

The above three on the other hand are much more intricately related by sheer logic.

Euclid spent decades to see these logical lines lying behind the plausible identity of the three.

This is highschool stuff but the three meanings should be introduced in Elementary School.

How far we go with some concepts is not determined by a didactical logic. In fact a didactical logic is not even concerned with the mentioned Logic of Actions. If this exists at all, it must be way above didactics. Didactics is the logic of understanding and it becomes independent of the actions. There seems to be a contradiction here because just earlier I emphasized the role of actions. Without doing math one can not become a mathematician. And yet understanding something new in math is the same for a mathematician as an outsider. In fact an outsider is a better measurer of didactical correctness. The actional background is not part of the understanding, it merely allows a wider application of it. Action is already involved in any understanding and can be pursued to widen itself. So it's only a very minimal fluency in actions that is really needed to even qualify to be a mathematician. After twelve years of math, everybody should become one. The cause of failure is the same as in all other subjects. The minimal fluency is intentionally avoided. A lot of bullshit is used to cover up the minimal necessity. A better practical system could be very easy. All education should contain three categories: The Roads, the Gardens and the Maps.

The road is the minimal "must". It can not be marked, it must be passed by the whole class!

This is what students must teach to each other too. In elementary school math, this is numeracy and solving word problems. Everything else then can be appreciation of gardens, that is problems raised by the teacher and active students. So learning as teaching appears on a higher level here. The maps tell the bigger pictures without even seeing the details. This is what proves in the most external way that the students are guided tourists and not cordoned animals.

Now back to fractions, there is the third mentioned application in the title, the three algorithms.

How this should be handled in elementary school that is as road or garden is up for discussion.

Observe that parallelity was Euclid's baby but we already mentioned his algorithm too, to fast track to the super common divider of two numbers n and d . We even showed how this is not only useful to simplify the $\frac{n}{d}$ fraction but also theoretically it proves the whole expansion

claim of the equal fractions. The logic was the same common expansion of the two fractions as the common multiples of two connectors to see parallelity. So Euclid's parallelity is hidden!

Euclid's algorithm can be used directly with fractions on three levels:

1. Dual algorithm:

If an $\frac{n}{d}$ fraction has a "smaller" variant, then either it is a simplification or they have an even

"smaller" common simplification. Let $\frac{n_1}{d_1}$ be the "smaller" variant and not being a

simplification. We can form a new $\frac{n_2}{d_2}$ fraction with n_2 being the remainder of n_1 in n

and d_2 the remainder of d_1 in d . If this $\frac{n_2}{d_2}$ is again not simplification of $\frac{n_1}{d_1}$ then we can

continue similarly. The end must be an $\frac{n_k}{d_k}$ which is a simplification of the previous, and thus

all earlier too, including the first two. For example: $\frac{72}{42} = \frac{48}{28}$. This is not a simplification, so

we can continue as $\frac{24}{14}$, which is a common simplification already.

2. Alternating algorithm.

The previous method only worked if we had already two variants to start from. A much better idea is to form again "smaller" fractions but now instead of being variants they have same simplifiers. To guarantee this is easy by changing the numerators or denominators to their remainders in each other. Of course again, repeat this change as many times as possible.

The final fraction must have zero remainder and so will have the super common divider of the original n and d in its numerator or denominator:

$$\frac{72}{42} \rightarrow \frac{30}{42} \rightarrow \frac{30}{12} \rightarrow \frac{6}{12} \rightarrow 6 \text{ is the super common divider of } 72 \text{ and } 42.$$

3. Mixed algorithm.

We can use mixed numbers to “calculate” these remainders above.

The alternation is simply a turning the fractional parts upside down:

$$\frac{72}{42} = 1 \frac{30}{42} \rightarrow \frac{42}{30} = 1 \frac{12}{30} \rightarrow \frac{30}{12} = 2 \frac{6}{12} \rightarrow \frac{12}{6} = 2$$

15. An unholy trinity, A useless grand result, A refinement, Fermat’s theorem

The title refers to three great concepts combined. Primes, factorial and dividability.

The atomness of primes at once means that arbitrary many numbers, all smaller than a p prime if multiplied together, will give a number not dividable by p . For composites the same is obviously not true because being composite, by definition means being the product of smaller numbers. So we might think that multiplying all smaller numbers than d means all possible compositions and therefore automatically being dividable by d . But this was a faulty logic because a composite can have repeated factors, so just multiplying all smaller numbers we used each only once. For example, $d = 4 = 2 \cdot 2$ and the smaller than 4 numbers multiplied is $(d-1)! = 3! = 2 \cdot 3 = 6$ and this is not dividable by 4. But actually only 4 is such composite. The next composite $d = 6 = 2 \cdot 3$ of course has no repeated factors, so the $(d-1)! = 5! = 2 \cdot 3 \cdot 4 \cdot 5 = 120$ is trivially dividable by 6. But then the next composite $d = 8 = 2 \cdot 2 \cdot 2$ has three repetitions of 2. Of course $(d-1)! = 7! = 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7$. There are three evens here, so all 2 factors of 8 are “covered”. This remains true because any f factor repeated r times means a $d \geq f^r$ number. The number of f multiples under this, is therefore at least $\frac{f^r - 1}{f}$ and this is always more than the repetition r , except for $f = r = 2$.

Indeed $\frac{f^r - 1}{f}$ is minimal at $f=2$ and $\frac{2^r - 1}{2} > r$ if $r > 2$ because at $r = 3$: $\frac{2^3 - 1}{2} > 3$.

So, for all $d > 4$ the $(d-1)!$ is always dividable by d if d is composite and never when d is a prime. Thus we got a simple test of primality. We don’t have to try all smaller numbers if they divide a d , rather calculate $(d-1)!$ and then check if d itself divides this.

This is the useless grand result in the title. Indeed, to calculate $(d-1)!$ is more work for large d values than checking the smaller numbers for dividabilities.

So now we can ask how the promised refinement will go. Well, it’s exactly how we refined the proof of the prime atomness. We went to multiples of a fix n start and to remainders.

Now with the factorial involved, this means calculating instead of $1 \cdot 2 \cdot 3 \cdot 4 \cdot \dots \cdot (d-1)$, the product $n \cdot 2n \cdot 3n \cdot \dots \cdot (d-1)n$ and checking the remainder of this. But beside this idea, we will also use the old refinement too, which was the fact that if n and d have no common dividers then all remainders appear for the multiples.

We’ll use again $[]$ for remainders and by the obvious logic of remainders remaining the same if only multiples of d are added, we can calculate the remainder of this big product in two alternative ways:

$$[n \cdot 2n \cdot \dots \cdot (d-1)n] = \begin{cases} [[n] [2n] \dots [(d-1)n]] = [(d-1)!] \\ [n \cdot \dots \cdot n \cdot 2 \cdot \dots \cdot d-1] = [n^{d-1} (d-1)!] = [[n^{d-1}] [(d-1)!]] \end{cases}$$

In the first version, we replaced the numbers at once to their remainders and then applied the mentioned old refinement result that all remainders appear. In the second version, we separated the n repetitions and then again replaced the members with their remainders.

Strangely, the first version's result, thus became one member in the second, and so we might say hastily that this at once shows, that either $[n^{d-1}] = 1$ or $[(d-1)!] = 0$.

So for $d = p$ primes then we at once get $[n^{d-1}] = 1$ because $[(p-1)!] \neq 0$.

Thus, indeed we obtained a refinement from the useless factorial prime criteria as a power criteria. This not only claims non dividability but a definite 1 remainder. So just as useless to verify primes but maybe important as theory. It is indeed a theoretical result and bears Fermat's name. Though he only rediscovered and never proved the forgotten fact known by ancient Chinese mathematicians. But our argument above was faulty in many ways!

First of all, the either / or is simply not true! Indeed, with $n = 1$ and $d = p = \text{prime}$, they are both true. Correcting it by saying simply "or" instead of "either / or" is not enough. We have to prove this "or" and it must rely on p being prime, because otherwise we can have $r = [q r]$ without 0 or 1. For example, at $d = 10$, $4 = [6 \bullet 4]$

So the precise claim is that for a p prime divider: $r = [q r] \rightarrow q = 1$ or $r = 0$.

First of all, if $r \neq 0$ then $q \neq 0$, so $q \geq 1$ and $q r \geq r$.

Then, $[q r] - r = 0$, so $[q r] - r = [q r - r] = [(q-1)r] = 0$ too. But for prime divider, non zero remainder's product has also non zero remainder, so $q-1$ or r must be zero.

Since $r = [(p-1)!] \neq 0$ thus indeed, $q-1 = [n^{d-1}] - 1 = 0$

16. Ancient mistakes and new blind spots, The pseudo primes

It's interesting to check out the only composite d that defies the $[(d-1)!] = 0$ rule, that is $d = 4$. Of course, $n = 1$ always gives $[n^{d-1}] = 1$, so we only have to check $n = 2, 3$.

Here $[2^{4-1}] = [8] = 0 \neq 1$ and $[3^{4-1}] = [27] = 3 \neq 1$.

This might suggest that for other composites too, the same will hold, that is $[n^{d-1}] = 1$ is not just a law of primes, but a test too, because it can't be true for composite d -s.

But this is easy to refute! First of all, if $n = d+1$ then n has 1 remainder, so all powers too.

If $n = d-1$ then the even powers are again all 1, so for all odd d : $[(d-1)^{d-1}] = 1$

The smallest composite odd number is $d = 9$ and then with $n = 8$: $[8^{9-1}] = [8^8] = 1$

The next one is $d = 15$ and so again with $n = 14$: $[14^{15-1}] = [14^{14}] = 1$.

Choosing an n and then to find a composite d is much harder and usually bigger too.

$n = 7$	first d is	$25 = 5 \bullet 5$	indeed	$[7^{25-1}] = [7^{24}] = [(7^2)^{12}] = [49^{12}] =$ $[(2 \bullet 25 - 1)^{12}] = [m \bullet 25 + 1] = 1$
$n = 6$		$35 = 5 \bullet 7$		$[6^{35-1}] = [6^{34}] = [(6^2)^{17}] = [36^{17}] =$ $[(35 + 1)^{17}] = [m \bullet 35 + 1] = 1$
$n = 5$		$124 = 2 \bullet 2 \bullet 31$		$[5^{124-1}] = [5^{123}] = [(5^3)^{41}] = [125^{41}] =$ $[(124 + 1)^{41}] = [m \bullet 124 + 1] = 1$
$n = 4$		$15 = 3 \bullet 5$		$[4^{15-1}] = [4^{14}] = [(4^2)^7] = [16^7] =$ $[(15 + 1)^7] = [m \bullet 15 + 1] = 1$
$n = 3$		$91 = 7 \bullet 13$		$[3^{91-1}] = [3^{90}] = [(3^6)^{15}] = [729^{15}] =$ $[(8 \bullet 91 + 1)^{15}] = [m \bullet 91 + 1] = 1$
$n = 2$		$341 = 11 \bullet 31$		$[2^{341-1}] = [2^{340}] = [(2^{10})^{34}] = [1024^{34}] =$ $[(3 \bullet 341 + 1)^{34}] = [m \bullet 341 + 1] = 1$

It is strange that the smallest $n = 2$ required the biggest $d = 341$.

An overall fact is that for any fix n , these composite d -s that satisfy $[n^{d-1}] = 1$ are much more rare than the primes that all satisfy it. So the new terminology is that all d -s that satisfy $[n^{d-1}] = 1$ for a fix n , are called n -probable primes. This includes of course, all primes and the non primes are called n -pseudo primes.

The largeness of the first 2-pseudo prime, 341 reflects that 2-pseudo primes are especially rare. In spite of all this, 341 is not a big number and the proof above for satisfying $[2^{341-1}] = 1$, was not difficult at all. So it's a mystery why this only happened in 1819 by F Sarrus.

The original explorers were ancient Chinese mathematicians. They especially liked to play with doubling numbers, so calculated the 2, 4, 8, 16, . . . sequence as far as they could. Then, subtracted or added 1 to these and examined them, especially for whether they are primes.

Today, due to our better notations (silver platters) we can just as well start from any n number and multiply it repeatedly, to get the n, n^2, n^3, \dots, n^d powers.

Then, $n^d - 1$ and $n^d + 1$ are the "power twins" and the main question is, if they are primes.

The only obvious fact is that if n is odd, then n^d is odd too, so both power twins are even and thus, can't be primes.

The basic two laws tell some necessary consequences if the twins would be primes:

1. $n^d - 1$ is prime $\rightarrow n = 2, d = \text{prime}$
2. $n^d + 1$ is prime $\rightarrow n = \text{even}, d = 2^k$

These are of course, for non trivial, that is $n, d > 1$ situations. To show them:

- 1.) If $n > 2$ were then $n^d - 1 = (n - 1)[n^{d-1} + n^{d-2} + \dots] = \text{composite}$
If $d = a b$ were then $n^d - 1 = (n^a)^b - 1 = (n^a - 1)[(n^a)^{b-1} + \dots + 1] = \text{composite}$
- 2.) n can't be odd, because as we said, then $n^d + 1$ is even.
 d can't have an odd divider, because if it had, that is $d = a b$ with $a = \text{odd}$ then,
 $n^d + 1 = n^{ab} + 1 = (n^a)^b + 1 = (n^a + 1)[(n^a)^{b-1} - (n^a)^{b-2} + \dots + 1] = \text{composite}$

An immediate consequence of 1.), 2.) is that if both power twins are primes, then $n = 2, d = 2$. And indeed, 3 and 5 are primes. But in no other case, can the power twins be both primes. These 1.), 2.) laws were known to the ancient Chinese mathematicians for the $n = 2$ special cases. In fact, they thought that the laws are true in reverse too, that is:

- 1.) $2^p - 1$ is always prime.
- 2.) $2^{2^k} + 1$ is always prime.

The first few cases are good in both, but the fifth and the sixth ones fail :

- 1.) $2^2 - 1 = 3, 2^3 - 1 = 7, 2^5 - 1 = 31, 2^7 - 1 = 127,$
but then: $2^{11} - 1 = 2047 = 23 \cdot 89.$
- 2.) $2^{2^0} + 1 = 3, 2^{2^1} + 1 = 5, 2^{2^2} + 1 = 17, 2^{2^3} + 1 = 257, 2^{2^4} + 1 = 65537$
but then: $2^{2^5} + 1 = 2^{32} + 1 = 4,294,967,297 = 641 \cdot 6,700,417.$
This was first observed by Euler.

Strangely, the continuation is very different for the two :

For $2^p - 1$, we do have both primes and composites re-appearing, but for $2^{2^k} + 1$, after the failing $k = 5$, no successful prime was found at all.

We don't have a proofs that these behaviors continue to infinity.

The $[n^{p-1}] = 1$ law was also known to the Chinese mathematicians for the special $[2^{p-1}] = 1$ case. But again, they believed in its false reversal.

When in Europe, Fermat, the French judge and amateur mathematician, rediscovered these old Chinese prime laws, he generalized them from doublings to powers, but still without proofs and questioning the reversals.

The first to prove $[n^{p-1}] = 1$ was Leibniz, but the name became Fermat's Little Theorem, because Fermat had a much more important discovery too, that became known as Fermat's Big or Last or Lost Theorem. After Leibniz, Euler gave an even better proof. Then Gauss generalized it to the extreme, and yet the reversal remained assumed.

Gauss missing the 341 composite case is almost beyond belief. I emphasize this, because it shows that the "blind spot" or "emperor's clothes" or "hypnotic spell" phenomenon is not only true for basic discoveries like Set Theory, but even for ongoing details. I do believe that we miss a lot of simple truths! But most importantly, I believe that a hypnotic spell as a veil covers our eyes. This particular case is very educational! Observe again the "easy" proof:

$$2^{341-1} = 2^{340} = 2^{10 \cdot 34} = (2^{10})^{34} = 1024^{34} = (3 \cdot 341 + 1)^{34} = M \cdot 341 + 1$$

Today, it is "obvious" that $(m d + 1)^k = M d + 1$ because we see that:

$(m d + 1) (m d + 1) \dots (m d + 1)$ is a sum containing all possible products, picking a member from each $(m d + 1)$. This will contain d in all members except the 1 resulting from the all 1 picks. It doesn't matter that we have $k = 34$ members or more.

But lets not forget that 2^{340} is a number bigger than all the atoms in the universe!

Only "empty" abstraction can conquer such number. So, here we are facing the advantages and evils of Formalism in its simplest world, namely among the natural numbers.

17. The square remainders

The extreme product of all non zero remainders, that is $(d - 1)!$ was successful as a theoretical primality test and also lead to the $[n^{d-1}] = 1$ Fermat's Little Theorem.

Strangely, the $[(p - 1)!]$ value itself didn't become determined. All we said is that it is non zero. Even more strangely, now we repeat this whole affair. Again, we'll use this $[(p - 1)!]$ without telling its value. But now, luckily it will come out by itself!

The basic fact from which we start is again the same, namely that $[r], [2r], \dots, [(p - 1)r]$ give all the non zero remainders. Here we used r instead of n and this formal difference reflects a new twist. Indeed, the same fact of all possible remainders among the multiples can also be said as follows: Choosing any q and r remainders from $1, 2, \dots, p - 1$ we can always find a second s pair to r , so that $[rs] = q$. Lets fix a q and go through the remainders as possible r values. Lets move the later s pairs backwards to follow the r -s :

1 , 2 , 3 , 4 , 5 , . . . , p - 1

1 , q , 2 , ? , 3 , ? , . . .

After 1 we can surely put q because $[1q] = [q] = q$.

Doing the same in the $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot \dots \cdot (p - 1)$ product:

$$[1 \cdot 2 \cdot 3 \cdot \dots \cdot (p - 1)] = [1 \cdot q \cdot 2 \cdot ? \cdot \dots] = \underbrace{[1q]}_q \underbrace{[2?]}_q \dots = [q^{\frac{p-1}{2}}]$$

So taking the remainders in the product as we did before, we obtained half as many, that is, $\frac{p-1}{2}$ many fix q members, and thus, at once obtained the factorial remainder value.

If it's that simple, then why didn't we give this before? Well because it's not that simple!

We made a big mistake! It is true, that every r has an s so that $[rs] = q$ but calling the s as "second" was false. Namely, s can be r itself!

So if the chosen q value is such, that $[r^2] = q$, that is q is a square remainder, then we'll encounter an r where this "bringing next to it" business wouldn't work.

In fact, if q is such square, then there are exactly two such "naughty" members, r and $p - r$.

Indeed: If $[r^2] = q$ then $[(p-r)^2] = [p^2 - 2pr + r^2] = [r^2] = q$ too.

And this $p-r$ is the only possible s second member, so that $[s^2] = a$, because:

$[r^2] = [s^2] = q$ with $r < s$ choice means that $r^2 - s^2 = mp$ and $r^2 - s^2 = (r-s)(r+s)$.

But, $r-s$ can't be dividable by p , so $r+s$ must be. But $r+s < 2p \rightarrow r+s = p \rightarrow s = p-r$

Now that we know that exactly these two are causing trouble for square q -s, we can make a new law for such square q -s by again pairing the members with the exception of $r, p-r$.

$$[1 \cdot 2 \cdot 3 \cdot \dots \cdot (p-1)] = [r(p-r) q^{\frac{p-1}{2}-1}] = [[rp - r^2] q^{\frac{p-1}{2}-1}] =$$

$$[(p - [r^2]) q^{\frac{p-1}{2}-1}] = [(p-q) q^{\frac{p-1}{2}-1}] = [p q^{\frac{p-1}{2}-1} - q^{\frac{p-1}{2}}] = p - [q^{\frac{p-1}{2}}]$$

We again used remainders in products, but also twice the fact that: $[mp - x] = p - [x]$

Thus, our combined result could be called shortly the Factorial Split or by a longer name:

Factorial Half Power Split By Squareness:

$$[(p-1)!] = \begin{cases} [q^{\frac{p-1}{2}}] & \text{if } q \neq [r^2] \\ p - [q^{\frac{p-1}{2}}] & \text{if } q = [r^2] \end{cases}$$

So indeed, as I promised, we obtained a law again, by the factorial without giving it exactly. But as I said too, now it will come out at once. Indeed, we do have an obvious square q value, namely $q = 1$. The roots of it are of course 1 and $p-1$. Indeed, $[1^2] = 1$ so the other is $p-1$.

So using $q = 1$ above: $[(p-1)!] = p - [1^{\frac{p-1}{2}}] = p-1$.

Finally we obtained a factorial's remainder value. This law, is called Wilson's Theorem.

Putting the obtained $p-1$ value back into the above Factorial Split Theorem:

$$p-1 = \begin{cases} [q^{\frac{p-1}{2}}] & \text{if } q \neq [r^2] \\ p - [q^{\frac{p-1}{2}}] & \text{if } q = [r^2] \end{cases}$$

The second thus means, $[q^{\frac{p-1}{2}}] = 1$ and so we obtained at once a new split:

Half Power Split or Half Power Split By Squares, or Euler's Criterion:

$$[q^{\frac{p-1}{2}}] = \begin{cases} p-1 & \text{if } q \neq [r^2] \\ 1 & \text{if } q = [r^2] \end{cases}$$

This actually is a refinement of Fermat's Little Theorem.

Indeed, knowing the half $\frac{p-1}{2}$ power, gives at once the full $p-1$ power as its square:

$$[q^{p-1}] = [(q^{\frac{p-1}{2}})^2] = \begin{cases} [(p-1)^2] = [p^2 - 2p + 1] = 1 \\ [1^2] = 1 \end{cases}$$

But there are much more important results that we can get from Euler's Criterion.

First we observe a new kind of split. This is simply whether the half $\frac{p-1}{2}$ is even or odd.

Indeed, p is prime, so always odd, if we ignore 2. So $p-1$ is even, and $\frac{p-1}{2}$ is a whole.

This itself can be even, or odd. Of course, $\frac{p-1}{2} = 2k$ means $p-1 = 4k$, that is $p = 4k + 1$.

While, $\frac{p-1}{2} = 2k-1$ means $p-1 = 4k-2$ that is, $p = 4k - 1$.

Looking through the primes, this $4k-1$, $4k+1$ split doesn't look particularly interesting. For example, circling the first ones and squaring the others, we find:

(3) [5] (7) (11) [13] [17] (19) (23) [29] (31) [37] [41] (43) (47) . . .

This seems quite random, just like the primes themselves among the naturals. Yet soon, we'll see an amazing exact law relating to these circles and squares. But right now, we should start from something obvious, regarding the evenness or oddness of $\frac{p-1}{2}$. This even or odd split at once interacts with another split like feature that we already encountered at the two square roots of q , if it was a square. As we showed, if one is r then the other is $s = p - r$. Such r , $p - r$ pairs can be called complementers or simply symmetrical pairs, which is visual too.

$$1 \quad 2 \quad . \quad . \quad r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad p-r \quad . \quad . \quad p-2 \quad p-1$$

Now we ask about these complementers or symmetrical pairs, not as potential roots of a fix a rather the q , $p - q$ themselves. Then obviously, it will involve whether a is square or not. By the way, since every square a has two distinct roots, thus exactly half of the remainders are squares and half not, so it is already a perfect situation for some symmetry law. But this will be a strange one, that itself is split, namely according to the parity of $\frac{p-1}{2}$, that is $p = 4k + 1$ or $p = 4k - 1$. But most amazingly, the road to discover this seemingly strictly square business is to examine the complementers in the half powers:

$$[(p-q)^{\frac{p-1}{2}}] = [m p + (-q)^{\frac{p-1}{2}}] = \begin{cases} [q^{\frac{p-1}{2}}] \text{ if } \frac{p-1}{2} = 2k \text{ so } p = 4k + 1 \\ p - [q^{\frac{p-1}{2}}] \text{ if } \frac{p-1}{2} = 2k - 1 \text{ so } p = 4k - 1 \end{cases}$$

Indeed, $(-q)$ to an even power is plus, while to an odd is minus.

These are exactly the values of the $[(p-1)!]$ split, but that was according to whether q is not

square or square. But, $[q^{\frac{p-1}{2}}] = p-1$ or 1 and these two can't be equal, so this identical split can only happen if the complements themselves obey two laws:

For $p = 4k + 1$, the square q -s are complements among themselves and thus of course, the non squares too.

For $p = 4k - 1$, the complements of square q -s are exactly the non square q -s.

$q = 1$ is always a square. Thus:

$p - 1$ is always a square for $p = 4k + 1$ but never for $p = 4k - 1$.

The $p = 4k + 1$ case can be expressed in another way too:

$$p = 4k + 1 \rightarrow p - 1 = [r^2] \rightarrow r^2 = m p + p - 1 = (m + 1)p - 1 \rightarrow r^2 + 1 = (m + 1)p$$

In short: Every $4k + 1$ prime divides an $r^2 + 1$.

This simple rule is the essence of this long journey. I tried to paint a bigger picture of the landscape, with the factorials, powers and beautiful world of square remainders.

This last single fact will be all we need from above to explore the $4k + 1$, $4k - 1$ duality of primes. Unfortunately, this fact that $4k + 1$ primes divide an $r^2 + 1$ is not enough. We have to make another journey, even more amazing than the world of remainders to a p prime.

18. Wilson Theorem again

Before we turn to this new journey, I want to show an interesting detour.

The obtained r remainder for which $r^2 + 1$ is dividable by the $4k + 1$ prime was merely an existence, so we may wonder if we can tell it exactly by a formula. Unfortunately, not!

But it will be enough for us that the $4k + 1$ primes divide any $n^2 + 1$ and we can give such n exactly too. This wouldn't be enough cause for this detour, but this exact n value is relating to the already mentioned Wilson Theorem. This was the fact that $[(p-1)!] = p - 1$.

In fact, we can get this result in a much easier way that also provides the exact $n^2 + 1$.

So starting from scratch, we can use the $q = 1$ product value only, that is accept the claim that for every r remainder, there is an s pair, so that $[rs] = 1$. Or to avoid even the use of remainders $rs = m p + 1$. Of course, again the 1 and $p - 1$ remainders would cause trouble, because they are their own pairs. But now, we simply omit these remainders, that is we only regard the product of the remainders from 2 to $p - 2$. Then:

$$2 \cdot 3 \cdot \dots \cdot (p-2) = \left(2 \cdot \frac{p+1}{2}\right) (3 \cdot ?) \dots = (p+1)(m p + 1) \dots = M p + 1$$

Obviously, there are $\frac{p-2-1}{2} = \frac{p-3}{2}$ many members, but luckily that's totally irrelevant, because the product is $M p + 1$ anyway. This at once means for remainders, that:

$$[2 \cdot 3 \cdot \dots \cdot (p-2)] = [(p-2)!] = 1$$

So amazingly, the factorial up to $p - 2$ is simpler than up to $p - 1$.

Up to $p - 1$ of course follows from the above by multiplying both sides with $p - 1$.

Now comes the even more amazing part that from the above $[(p-2)!] = 1$ version, we can get a new even much smaller factorial version, namely only up to $\frac{p-1}{2}$.

We can do this by writing the second half of the product from 2 to $p - 2$ in reverse order:

$$2 \cdot 3 \cdot \dots \cdot \frac{p-1}{2} \cdot \frac{p+1}{2} \cdot \dots \cdot (p-2) = 2 \cdot 3 \cdot \dots \cdot \frac{p-1}{2} \cdot (p-2) \cdot \dots \cdot \left(p - \frac{p-1}{2}\right) =$$

$$2 \cdot 3 \cdot \dots \cdot \frac{p-1}{2} \left[mp + (-1)^{\left(\frac{p-1}{2}-1\right)} \cdot 2 \cdot 3 \cdot \dots \cdot \frac{p-1}{2} \right] = N p + (-1)^{\left(\frac{p-1}{2}-1\right)} \left(\frac{p-1}{2} ! \right)^2 =$$

$$= (p-2)! = M p + 1$$

If $p = 4k + 1$ then $\frac{p-1}{2} - 1 = 2k - 1 = \text{odd}$ so:

$$-\left(\frac{p-1}{2} !\right)^2 = M p + 1 - N p \quad \text{from which} \quad \left(\frac{p-1}{2} !\right)^2 + 1 = (N - M) p$$

If $p = 4k - 1$ then $\frac{p-1}{2} - 1 = 2k - 2 = \text{even}$ so:

$$\left(\frac{p-1}{2} !\right)^2 = M p + 1 - N p \quad \text{from which} \quad \left(\frac{p-1}{2} !\right)^2 - 1 = (M - N) p$$

19. Square sums and primes

I don't want to be a tease, so I reveal the law behind the $4k + 1$, $4k - 1$ split of primes. The $4k + 1$ primes are square sums, namely only one way, while the $4k - 1$ primes, never:

$$3 \neq a^2 + b^2, \quad 5 = 1^2 + 2^2, \quad 7 \neq a^2 + b^2, \quad 11 \neq a^2 + b^2, \quad 13 = 2^2 + 3^2$$

$$17 = 1^2 + 4^2, \quad 19 \neq a^2 + b^2, \quad 23 \neq a^2 + b^2, \quad 29 = 2^2 + 5^2, \quad 31 \neq a^2 + b^2, \dots$$

Amazingly, one half of our claim is almost self evident, namely an even more general fact is:

No $4k - 1$ number can be square or square sum ! Indeed:

- 1.) Even number's square is $(2k)^2 = 4k^2 = 4m$
- 2.) Odd number's square is $(2k+1)^2 = 4k^2 + 4k + 1 = 4n + 1$
- 3.) From 1.) and 2.) we see at once that a square can't be $4k - 1$.
- 4.) If a, b are both even, then $a^2 + b^2 = \text{even} + \text{even} = \text{even}$.
- 5.) If a, b are both odd, then $a^2 + b^2 = \text{odd} + \text{odd} = \text{even}$.
- 6.) If one of a, b is even and the other is odd, then $a^2 + b^2 = 4m + 4n + 1 = 4k + 1$.
- 7.) So as we see $a^2 + b^2$ can never be $4k - 1$ either.

Thus, one half of the splitting of primes is solved. The $4k - 1$ ones can't be square sums. Strangely, the uniqueness of $4k + 1$ primes as square sums is again a simple and wider fact:

If an n number is square or square sum in two ways, then n is composite! Indeed:

First of all $n > 2$ because 2 only has one form as $1 + 1$.

All $n > 2$ evens are composite, so enough to prove it for odd n .

Thus, in a square sum form, the two squares must be different, that is:

$$n = A^2 + a^2 = B^2 + b^2 \quad \text{with} \quad A > a \geq 0 \quad \text{and} \quad B > b \geq 0 \quad \text{Then:} \quad n [A^2 - a^2 + B^2 - b^2] =$$

$$n (A^2 - a^2) + n (B^2 - b^2) = (B^2 + b^2)(A^2 - a^2) + (A^2 + a^2)(B^2 - b^2) =$$

$$2(A^2 B^2 - a^2 b^2) = 2(AB - ab)(AB + ab) \quad \text{and lets observe that:}$$

$$AB - a^2b < AB + a^2b = \frac{A^2 + B^2 - (A - B)^2}{2} + \frac{a^2 + b^2 - (a - b)^2}{2} < \frac{A^2 + a^2 + B^2 + b^2}{2} = n$$

So $n \in \mathbb{N} = 2 \cdot (\) \cdot (\)$ with both $(\)$ numbers $< n$.

But a prime can't divide a product of smaller numbers, so n must be composite.

By the way, such composites do exist, for example,

$$25 = 5^2 = 3^2 + 4^2, \quad 65 = 1^2 + 8^2 = 4^2 + 7^2, \quad 85 = 2^2 + 9^2 = 6^2 + 7^2$$

Thus, the only thing remains to prove is that all $4k + 1$ primes are square sums. But, we ended the previous section as final result, that all $4k + 1$ primes divide an $r^2 + 1$.

This is a special square sum because $1 = 1^2$, but that's too special and hides the real importance of the specialness. This real importance for any $a^2 + b^2$ square sum is if a and b have no common divider except 1. Indeed, square sums behave a bit like fractions.

If A, B have a c common divider, then $A^2 + B^2 = (ca)^2 + (cb)^2 = c^2(a^2 + b^2)$, so we have a simpler $a^2 + b^2$ already. If c was the super divider of A, B then $a^2 + b^2$ is a simple form, a, b have no common divider anymore.

Now, the result of the last section means, that all $4k + 1$ primes divide a simple square sum.

Thus, it would be enough to prove that:

All prime factors of a simple square sum are square sums.

Amazingly, again much more is true, regardless of primes:

All non 1 dividers of a simple square sum are also square sums.

Of course, these dividers then must be simple themselves, because a non simple square sum has a c^2 divider and this would divide the original number too, which then couldn't be simple.

So the simple square sums are numbers that form their own world as far as division is concerned! Indeed, lets check them out:

$$2 = 1^2 + 1^2, \quad 5 = 1^2 + 2^2, \quad 10 = 1^2 + 3^2, \quad 13 = 2^2 + 3^2, \quad 17 = 1^2 + 4^2$$

$$25 = 3^2 + 4^2, \quad 29 = 2^2 + 5^2, \quad 37 = 1^2 + 6^2, \quad 41 = 4^2 + 5^2, \quad 50 = 1^2 + 7^2$$

$$53 = 2^2 + 7^2, \quad 61 = 5^2 + 6^2, \quad 65 = 1^2 + 8^2 = 4^2 + 7^2, \quad \dots$$

I went up to 65 because it has two forms, so it shows that unlike just among the primes, uniqueness is not guaranteed in general. By the way, the two square sum forms of 65 are not apparent from the square sum forms of its prime factors, 5 and 13. What's even stranger, is that 10 and 50 were earlier composites but these have only one form.

Going further in the examples we can observe that all new ones are either primes or a product of earlier ones. But not all product of earlier become new ones. Namely, odds can be freely multiplied, but even ones, only with odds. This also means:

A number is simple square sum, if and only if, its prime factorization contains only square sum primes, but the 2 maximum once.

This at once implies the previous rule, that all dividers of the simple square sums are themselves simple square sums. Also, to prove this perfect characterization, it would be enough to prove our first rule, that is, that the prime factors of simple square sums are already such, plus the "if" part of the characterization, that is the inheritance to prime products, containing only one 2.

The single allowance of 2 seems to explain why the evens are special and so the 10 and 50 had only one forms. But we are wrong! The double of 65 again has two forms, and what's worse totally unrelated to the ones 65 had : $130 = 3^2 + 11^2 = 7^2 + 9^2$.

We might also contemplate about the repeating of a 2 factor as being dividable by $4 = 2^2$ and thus causing the 2 in both members of the square sum, so stopping simplicity.

Indeed $2^2 \cdot 5 = 20 = 2^2 + 4^2 = 2^2 (1^2 + 2^2)$ is a non simple square sum. But this logic is false too! First of all, it merely means that a non simple square sum form exists and second, it applies to any other repeating factor too. For example instead of $2^2 \cdot 5 = 20$ we can take $2 \cdot 5^2 = 50$. And indeed, this has the $25 + 25$ non simple square sum form. But here other simple form does exist $50 = 1^2 + 7^2$, so 50 is a simple square sum.

In spite of this false logic it's amazingly easy to show that the single appearance of 2 must be true. Namely:

An $a^2 + b^2$ simple square sum can not be dividable by 4. Indeed:

a, b can't be both even by simplicity. If one is odd the other even then $a^2 + b^2$ is odd. If both are odd then $a^2 + b^2 = (2m + 1)^2 + (2n + 1)^2 = 4m^2 + 4m + 1 + 4n^2 + 4n + 1 = 4k + 2$.

In spite of this simple proof, the fact still begs the question:

Why does the "perfect" world of simple square sums contain this weird single usage of only 2? And how do the new square sum forms come about from the earlier ones?

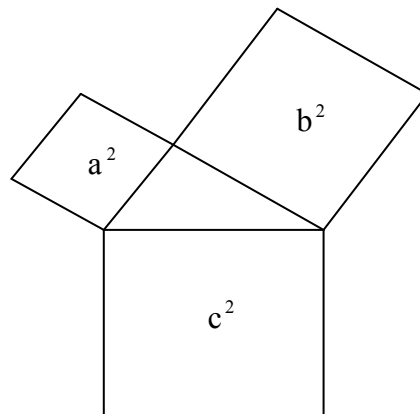
Gauss was the first who answered these questions and discovered the world where our journey will lead. But now, I want to prepare that journey with a strange coincidence.

Indeed, Gauss' discovery ventured into the plane and the most important law of plane geometry is the Pythagoras Theorem. But Pythagorean number triplets are also related to the prime splittings and square sums.

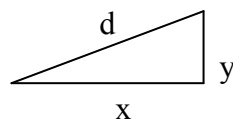
20. Pythagoras theorem: The greatest law

After Einstein discovered Special Relativity, it turned out that the usage of time as fourth coordinate, fits into a new four dimensional Pythagoras Theorem.

The three dimensional version is merely a repeated application of the simple plane geometrical law of $a^2 + b^2 = c^2$.

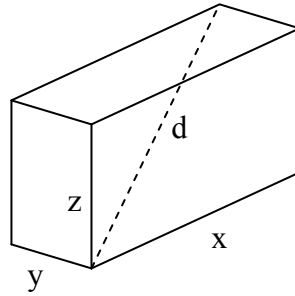


This old fashioned visualization is actually much less useful than a more modern coordinate version:

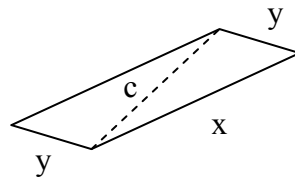


$$d^2 = x^2 + y^2$$

Then we can visualize a d distance in space between two points that have x, y, z coordinates. This is merely a brick with x, y, z edges and we can calculate the d diagonal:

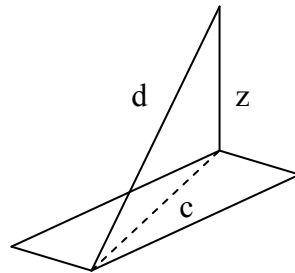


First we can calculate the c bottom diagonal with one usage of the Pythagoras Theorem:



$$c^2 = x^2 + y^2$$

Then a new triangle appears with c and z as sides giving the d diagonal as the second usage of the Pythagoras Theorem.

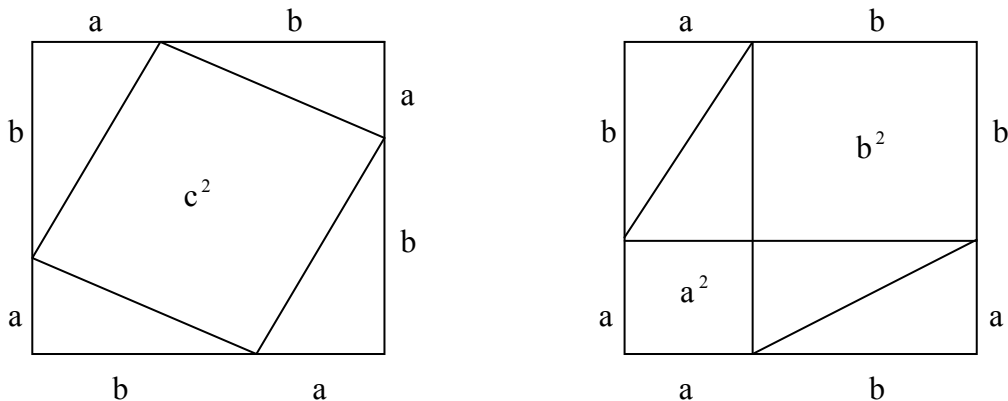


$$d^2 = c^2 + z^2$$

Replacing $x^2 + y^2$ into c^2 we obtain $d^2 = x^2 + y^2 + z^2$.

This is what Einstein continued with a strange time member to get the “event distance”.

Indian mathematicians proved the Pythagoras Theorem first with the Hindu method, where speech was not allowed. The pictures had to be so convincing, that the nodding of the listener was enough. The picture they used was the following:

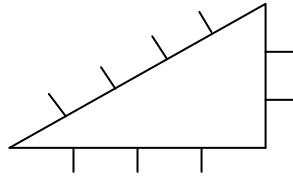


Both squares are $a + b$ sided ones and contain four of the triangles. The leftover in the first is c^2 , while in the second, $a^2 + b^2$. So indeed, these have to be equal.

21. No common units, The Geometrical Euclidian Algorithm

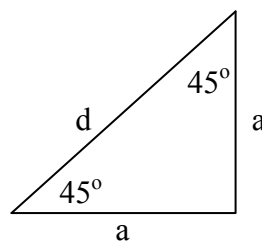
The ancient geometers were especially interested in those rare right angled triangles where all three sides are multiples of a common unit.

The simplest is $3^2 + 4^2 = 5^2$



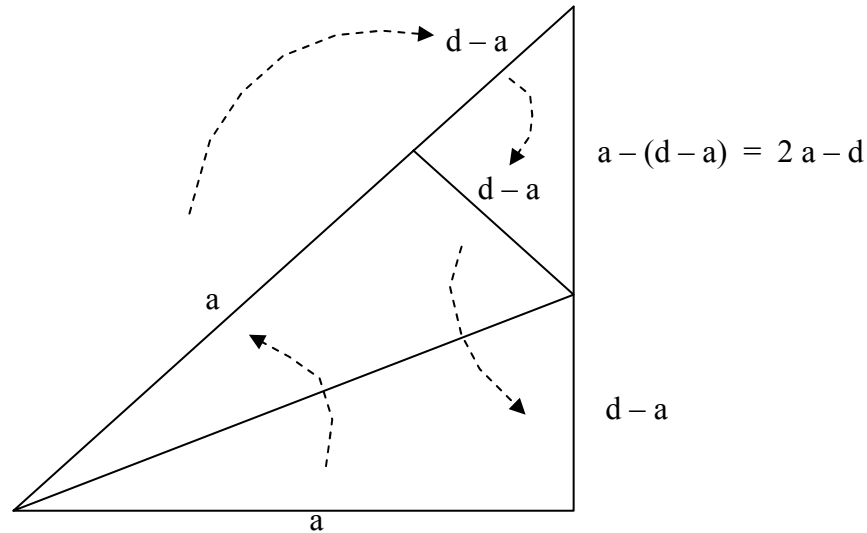
This is understandable because they were obsessed in general with the problem of measuring different distances with same units. The infinite grid system of Gauss would have been the perfect background for this and yet they didn't pursue that. The picture above in the grid system, simply means that the $(4, 3)$ connector's length is 5. This is very strange because the connectors don't have the tiles along and so shouldn't be whole values. An idea to force the wholeness could be to change the units to smaller and thus maybe fit that into a distance. This is what we do when we measure lengths with a tape measure. If the length is not an exact centimetre we simply use millimetres too, which means a whole number in these. Our infinite decimal system is the silver platter that makes it quite plausible that no matter how fine units we use, there will be distances that don't fit in because they are infinite decimals. For a while the Greek mathematicians actually believed that all distances can be measured in common units. But then without the infinite decimals, rather by individual tricks they realized too, that for any chosen a distance, there are many second d so that a and d can not have common units.

One ingenious method to prove this was used by Euclid. His algorithm we already mentioned for finding the super common divider of two numbers. His geometrical algorithm shows that he was almost aware of the concept of linear combinations. Indeed here with starting two distances instead of two numbers, he relaxed the remainders to merely smaller combinations. But the second more important difference is that here the goal is not finding a final minimal distance, rather quite oppositely use the fact that now the distances will diminish for ever to prove that there can be no common unit. Using differences, that is smaller linear combinations, the common units remain. But of course the common units are merely hypothetical, we want to prove that they don't exist. The decreasing of the combinations is then indeed proves this because for any fix unit length we can go under the unit length and create a distance that can't be measured in that unit. But there are two problems! Firstly, how do we define a potentially infinite sequence of combinations? And second we have to avoid a zero difference that is perfect fit which would exactly mean existing common unit. The solution to both problem is the same. We make the sequence cyclic and thus determined by merely finite members. A perfect cyclicity of course is impossible, because the distances must decrease. The compromise, is the essence of the plane, the similarity of proportionally altered objects, namely triangles. Observe that the other essence of the plane is shiftability which was already used in the concept of the linear combinations. The simplest cycle is to create from a triangle at once a smaller similar one, with sides being combinations of the original. Then that new one will create a new again and again. The simplest connector (a, a) already has a d distance that can not have common unit with a:



Euclid's method for this fact needs a cycle that creates a smaller similar triangle with sides being linear combinations of a and d.

The cycle starts with a mere subtraction of a from d , which can be achieved by mirroring the bottom a side to the angle halver line:



Then, the dotted arrows show the logic of derivation, how a smaller similar triangle appeared with $d - a$ sides corresponding to a and $2a - d$ to d .

To repeat again the logic that contradicts a common unit:

If a had a unit that fits into d too, then it would fit into $d - a$ and $2a - d$ too.

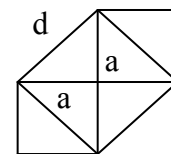
In other words, the little triangle would inherit the unit. But, in that, we can repeat our construction to obtain an even smaller version which would inherit the unit again, and so on.

Clearly, a unit couldn't fit into all those smaller and smaller distances that shrink into a point.

Knowing the algebraic relations of sides can create algebraic proofs for having no common units too. For example above, $a^2 + a^2 = 2a^2 = d^2$. So not having common unit, simply means that there are no m, n natural numbers, so that $2m^2 = n^2$.

And indeed, m^2 has either no 2 as prime factor or has even many. Thus, $2m^2$ has odd many. But, n^2 has also none or even many. So the two sides can't be equal.

By the way, we don't need the Pythagoras Theorem to see that $2a^2 = d^2$:



22. Pythagorean triplets

As I mentioned, the 5 length of the $(4, 3)$ connector is a rare coincidence allowed by the Pythagoras Theorem and by the algebraic coincidence that $3^2 + 4^2 = 5^2$.

This is how the Greeks and even the earlier Babylonians regarded it, that is more as an algebraic coincidence than a geometrical. This is very strange! Why did it take two thousand years for finally to go back to the geometrical roots by Gauss. Most sadly even after that, it all sank back to obscurity, instead of altering the way we look at the plane. We'll come to those beautiful visions, but now lets look at the "ugly" things.

The Babylonians practically knew all the possible Pythagorean triplets, like $3^2 + 4^2 = 5^2$.

Earlier we investigated the square sums in general without proofs. These Pythagorean triplets are merely special square sums where the sum is itself a square. So why didn't the Babylonians investigated the square sums in general? Probably the geometrical meaning was important for them. So as we see they fell in between geometry and algebra. Too narrow geometrical and too narrow algebraic view was that made their results ugly. The essence of their wisdom was that knowing how a c number is square sum, can easily tell how c^2 will be too:

If $c = u^2 + v^2$ with $u < v$ then, $c^2 = a^2 + b^2$ with $a = v^2 - u^2$ and $b = 2uv$. Indeed,
 $a^2 + b^2 = (v^2 - u^2)^2 + (2uv)^2 = v^4 + u^4 - 2v^2u^2 + 4u^2v^2 = u^4 + v^4 + 2u^2v^2 =$
 $(u^2 + v^2)^2 = c^2$

They obviously knew, that the reverse is not true because the first counter example is quite small, namely 15 is not a square sum and yet $15^2 = 225 = 81 + 144 = 9^2 + 12^2$.

Probably they also realized that this is not a simple square sum though because it can be simplified with 3^2 . So they merely believed in that the formulas give all simple triplets.

Knowing the law, we already revealed that all non 1 dividers of a simple square sum are also such, it's obvious that if c^2 is simple square sum, then c is too. In fact, from the mentioned bigger picture, that simple square sums contain only one 2 factor, even the more precise rule follows:

c^2 is simple square sum, if and only if, c is an odd simple square sum.

The big picture will come out from the journey we'll do with connectors, starting from the next section. But this can be shown directly, without the journey. Indeed:

For the "if" part we only have to show that the above used Babylonian $c^2 = (u^2 + v^2)^2 = (v^2 - u^2)^2 + (2uv)^2$ formula is simple if $c = u^2 + v^2$ is simple and odd.

It gives an odd because odd's square is odd. For the simpleness, due to the oddness it's enough to show that negatively, if a $p > 2$ prime would divide both $v^2 - u^2$ and $2uv$ then it would also divide u and v . And indeed p would divide $c^2 = (u^2 + v^2)^2$ so, $c = u^2 + v^2$ too and so $(u^2 + v^2) \pm (v^2 - u^2) = 2v, 2u$ too.

For the "only if" part, let $c^2 = a^2 + b^2$ with a, b having no common divider except 1. Then:

1.) a, b can't be both even.

2.) a, b can't be both odd, $2m + 1, 2n + 1$ because then
 $a^2 + b^2 = (2m + 1)^2 + (2n + 1)^2 = 4m^2 + 4m + 1 + 4n^2 + 4n + 1 = 4k + 2$
 couldn't be a square. Indeed, it's even and an even square is always $4k$.

3.) So exactly one of a, b is odd say a . Thus also a^2 is odd and b^2 is even. Then,
 $c^2 = a^2 + b^2$ is odd and thus c too. So $c - a, c + a$ are both even, and so
 $\frac{c-a}{2}, \frac{c+a}{2}$ are wholes.

4.) Also observe that $c^2 - a^2 = (c - a)(c + a) = b^2 = \text{even square} = (2k)^2 = 4k^2$. So
 $k^2 = \frac{c^2 - a^2}{4} = \frac{c-a}{2} \frac{c+a}{2}$ is a square too.

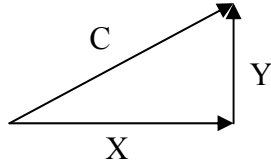
Now, $c - a$ and $c + a$ can't have common prime factors, because otherwise it would be common for a, c too and thus for a, b too. So $\frac{c-a}{2}$ and $\frac{c+a}{2}$ are both squares.

5.) Thus, we can take the square roots and $\sqrt{\frac{c-a}{2}} = u, \sqrt{\frac{c+a}{2}} = v$ will give the reverse.

$$\text{Indeed } u^2 + v^2 = \left(\sqrt{\frac{c-a}{2}}\right)^2 + \left(\sqrt{\frac{c+a}{2}}\right)^2 = \frac{c-a}{2} + \frac{c+a}{2} = c.$$

23. Connector Arithmetic

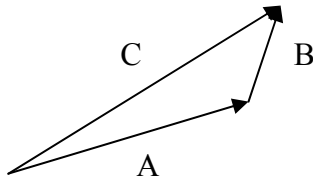
Both addition and multiplication of connectors are generalization of the coordinate concept. We already called the coordinates as components when regarded not simply as numbers rather as special horizontal and vertical connectors:



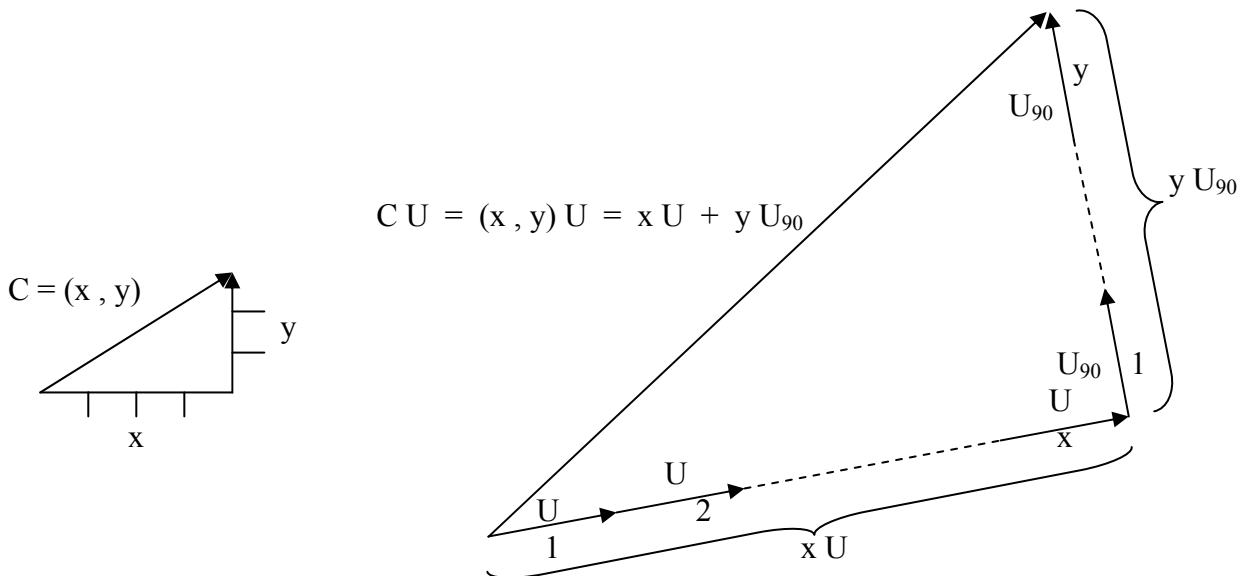
So then the C connector can be also regarded as the sum of these components: $C = X + Y$. We expect that a sum should be exchangeable in its order and indeed $X + Y$ and $Y + X$ are parallel same long and directional that is same too:



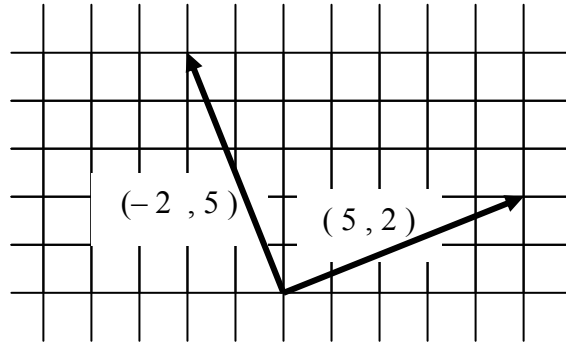
The generalization for arbitrary two A, B connectors is easy. We continue them. So in fact we generalized the original concept of the trip.



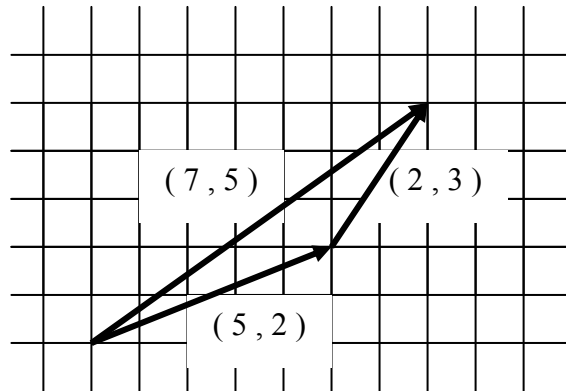
The real tricky concept that Gauss discovered was multiplication. We already used k-multiples of connectors for our heuristic proof of common multiples implying parallelity. The k-multiples are a kind of stretching of the $C = (x, y)$ connector into $kC = (kx, ky)$. A new idea is to alter the C connector not from the outside rather from the inside. So we go back to basics that is the trip again. To go x horizontally and then y vertically, can be applied to any U connector regarded as a new imaginary horizontal unit. Indeed we simply go in U steps x many times and then make a 90 degree turn and use the same unit y many times. In fact using addition as continuation allows the use of the old k-multiple to define this new inside one, if we use U and its 90 degree turned U_{90} too:



Observe that a 90 degree turn simply means using $-y$ for x and x for y :

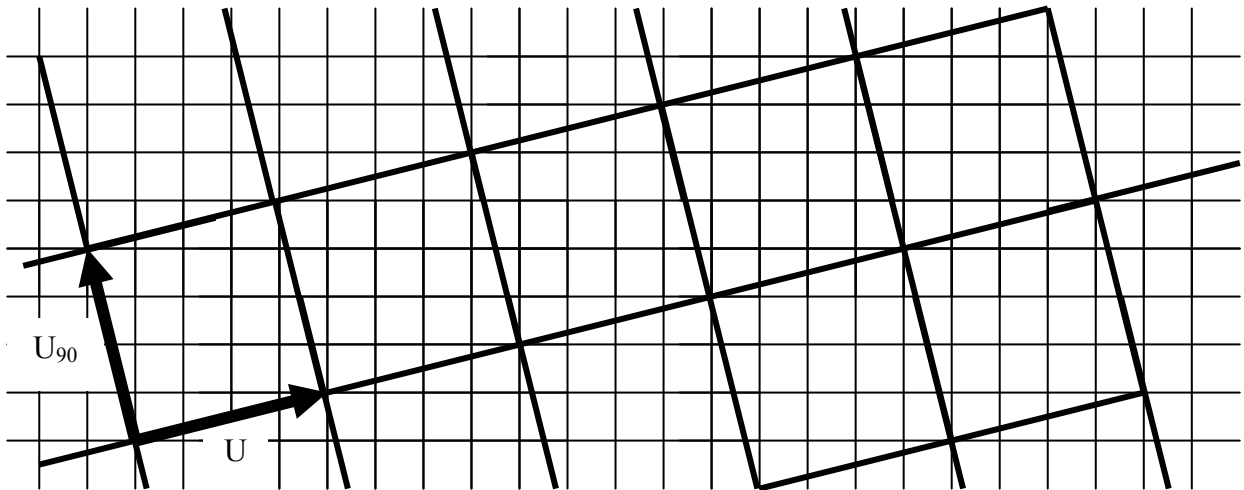


And that an addition of connectors means addition of coordinates:



Thus we can continue the calculation of products as: $C U = (x, y) U = x U + y U_{90} = x(u, v) + y(-v, u) = (x u, x v) + (-y v, y u) = (x u - y v, x v + y u)$.

But this fast tracking to the final multiplication from coordinates is missing the beauty of the original idea! Indeed, any U connector as new unit should be regarded as generating a new sub grid system inside the original:



So the $C U$ “inside” multiplication simply means that the first member C is placed into the new sub grid system generated by the second member U .

In fact, if we observe a trivial special choice of U , then this order of the inside U standing as second, is even more logical. Indeed, if U is chosen as $(1, 0)$ that is the horizontal unit connector, then the sub grid system is actually the full. So we have this trivial multiplication already as $C = C(1, 0)$. The $C U$ generalization is merely using U instead of $(1, 0)$.

But now comes the problem! Just as addition, multiplication is also exchangeable in its order. For addition too $3 + 4 = 4 + 3$ is actually a surprise if we start from counting. Indeed, why should we get to the same result by continuing counting from 3 extra 4 steps as continuing from 4 with 3 steps? The only thing that makes it trivial is by placing them into space! The one dimensional line is enough! The incorrect way is to use coordination, that is fix an arbitrary origin and represent the numbers as dots. The correct way is to chose a fix step on the line! The step is already part of the counting process but a fix start is not! So it's not some points of the line that we must order to the numbers rather the much wider, step from one point to an other! This includes two choices! The length of the step and its direction as to the right or to the left! As tradition, we can chose the right! Then from any point this step can be applied. Addition is the continuation of counting and placed into this reality of the line the sums become distances measured next to each other. The shiftability was already used as the representation of counting startable from any point. But then, even from a fix starting point we can apply the shiftability to see that $3 + 4 = 4 + 3$. Simply shift the 3 after the 4! So again, the representation of counting should be a fixed arrow to the right!!!!

The application must leave the freedom of start since the line is shiftable. In fact the shiftability and proportionality of space is used hidden already, because the blackboards are all over the world and different in size!!!!

Just as addition effectively determined over the spaceless abstract counting, necessitates the one dimensional line, the multiplication determined over addition necessitates the two dimensional plane! So as I mentioned $3 \bullet 4 = 4 \bullet 3$ is only evident as areas of the same rectangle. But we jumped a lot ahead because the fact that addition necessitates the line also means that it necessitates the reverse left directional arrow and thus the subtraction and the negative numbers and zero. So these can and should be visualized before the rectangles! But there is much more to all this! Remember I said that the "why-s" should only begin in high school. This doesn't mean that the merely visualized explanations don't have a logic too.

The times table is learnt before we draw the lines for addition. So there is a period of no visualization at all! Then the line of addition already necessitate the plane subconsciously! This is resolved by allowing vertical lines for additions with an upward directional step as counting. This alternative number line is much better for introducing the negative numbers and zero as sea-level.

The philosophy of mathematics includes these didactical logics. If it turns out to be true what I believe, that all didactical logic is mathematical then mathematics will take over philosophy! But until then philosophy must be respected as an outside field!

Newton said "Physics beware of metaphysics". This is a correct reflection on something that stopped physics before. But he didn't reflect on something much more important, namely how far is physics beyond mathematics. The bringing in of math was Kepler's grand act. But he falsely believed that reality is directly mathematical. The mediator of math into reality, that is physics, was discovered by Newton. The mystery of math, why is it true, how can we discover it and how can we all learn it was ignored before. Now that its reality over matter became expressed, it gained an instant respect. But mathematics went on uninterrupted. A new level of exactness was pursued by Gauss the greatest admirer of Newton. He not only avoided the same question, what is in physics beyond math, but also the question of his crucial obsession, what constitutes a mathematical proof and what lies beyond proof. And these still only relate to the truth of math and the other two, the discovering and understanding is not even mentioned. But he was more specific about philosophy and said that it either tells trivialities or false generalizations. It's amazing how such a genius could have been that narrow minded. Indeed, his contemporary countryman Hegel, happened to be the greatest philosopher up to the time just as he was the greatest mathematician. But what's more, Hegel did reply to Newton's motto "physics beware of metaphysics" by "So Newton actually said physics beware of thinking". A mathematician must respect at least someone who goes straight to the point. And this was neither a trivial nor a false reply. All true mathematicians are in continuous compensation because they already got the passport to heaven. All true philosophers have delusions of grandeur because they already got the visa to heaven.

Now we really return to the problem that the connector multiplication should be exchangeable. A Formalist would say “What’s the problem? Look at the already obtained coordinate form and it’s obviously exchangeable”. Indeed exchanging x and u also y and v in:

$(x u - y v, x v + y u)$ we get exactly the same if we already accept that:

$x u = u x, y v = v y, x v = v x, y u = u y$ and $x v + y u = y u + x v$.

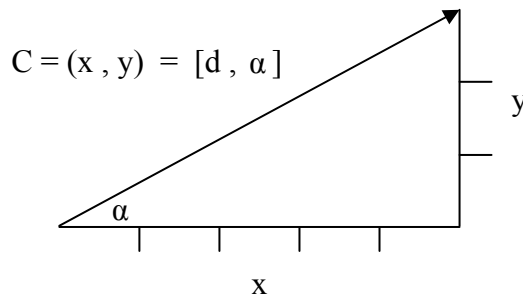
But this was not a visible explanation. Well, is there a visible one too? Of course there is!

This in fact necessitates the more general multiplication of plane or complex numbers that are not merely connectors of grid points but arbitrary points. And if you thought I won’t utter the word “necessitate” again then here is a fact from the history of philosophy: Necessitation was a crucial concept of Hegel’s dialectical evolution of thoughts. This doesn’t diminish the fact that he was an ignorant prick in relation to mathematics. He simply never practiced math, so he misjudged math similarly as Gauss misjudged philosophy, that it is regarding it merely as a chain of trivialities or false, namely useless generalizations.

Now lets turn to the philosophically correct new meaning of connector products:

Both the C and the U connectors must be regarded in a new way to see this new meaning.

Instead of the trips as horizontal and then vertical moves, we can regard them as instant move with a given angle and length:



So $[d, \alpha]$ denotes the d long and α angled connector. The angle is of course measured from the horizontal line anti clockwise as a tradition.

We can write the other U connector similarly: $U = (u, v) = [r, \beta]$

And now, the simple fact is: $C U = [d, \alpha][r, \beta] = [d r, \alpha + \beta]$

Indeed, we use r as unit length so the d length stretches to $d r$. And we use U as imaginary horizontal so the α angle is measured from U 's β .

From this new $[]$ coordination calculation we can derive the exchangeability didactically correctly because it carries the meaning too: Lengths multiply and angles add up. So if addition and multiplication of real numbers that is infinite decimals is exchangeable then connector multiplication is too.

We can't pretend that this derivation is a perfect affair because it still required the multiplication of real numbers. So we did go out of the grids. That multiplication is instantly visible as rectangle areas and its exchangeability follows from that too. But that was already needed for the simple $3 \bullet 4 = 4 \bullet 3$ natural case too! Rectangles can not be avoided to be didactically correct. And yet they have to be avoided! So here is an other Hegelian truth for the Formalists, namely the resolution of contradictings, or as he called it, the thesis and antithesis becoming synthesis. Once we establish that the multiplication of real lengths is exchangeable, we shouldn't drag in areas every time we use multiplication and this exchangeability.

This also means that unless we use the Pythagoras theorem as merely a fact about squares as geometrical objects we must regard the old Hindu proof as didactically incorrect too!

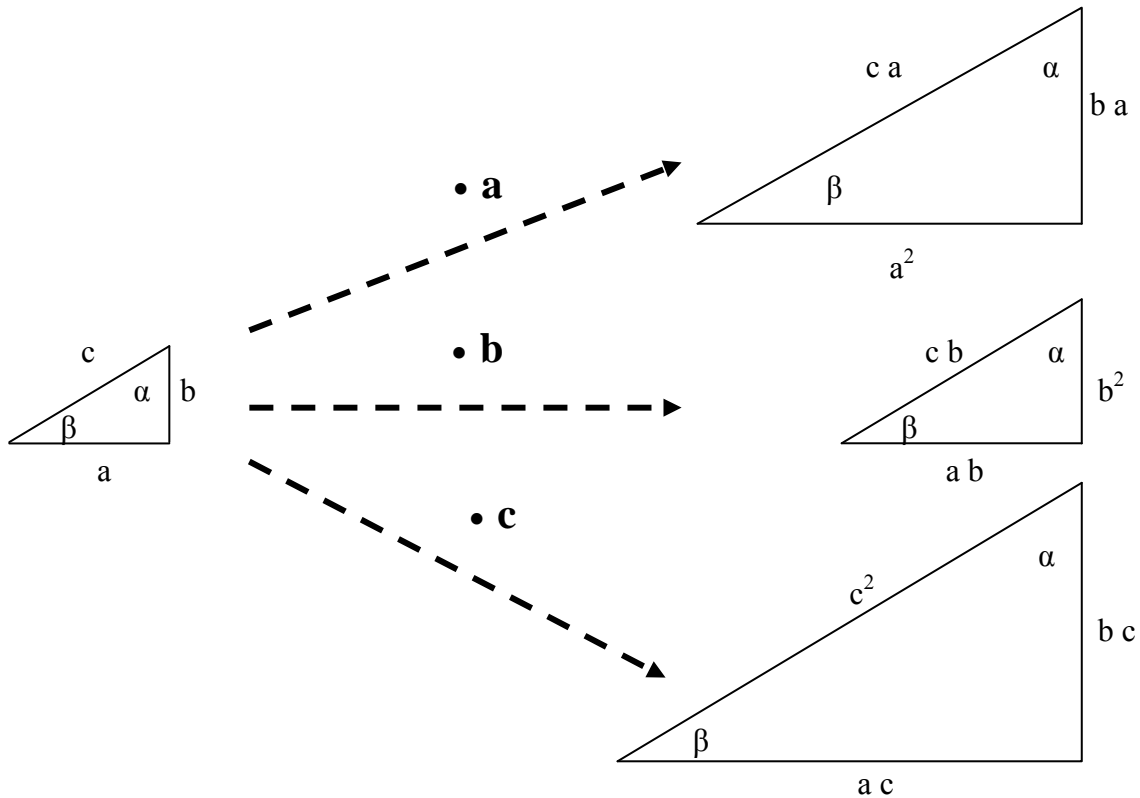
And of course we can not regard the Pythagoras Theorem as merely that. It tells the length of a connector from the coordinates. That it only tells the square of the length is not accidental.

So the Pythagoras Theorem “necessitates” the whole field of connectors and should be introduced as such!

24. Pythagoras Theorem, proved as it should be

We digress from the connectors for a “second” to put this most important theorem on correct grounds. This doesn’t mean that we have to swallow it up, that is incorporate it into the connectors. In fact I will start with a proof using the traditional $a^2 + b^2 = c^2$ notation, so not even mention connectors or coordinates. A perfect proof leads the student to the proof, so instead of the passive Hindu method of “watching”, we’ll create “action”.

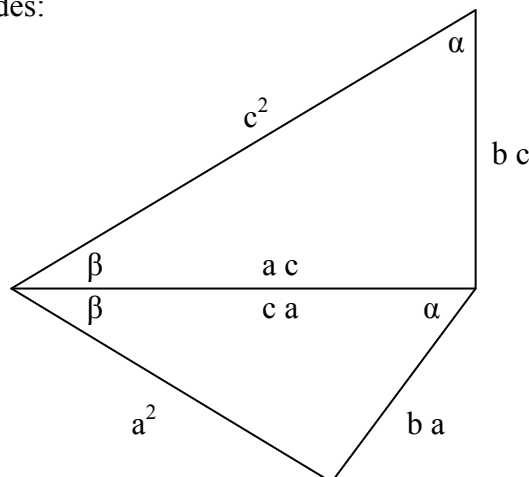
Since the claim contains all three sides multiplied with itself thus we ask the student to enlarge the a, b, c triangle in three ways namely multiplying every sides with a then b then c :



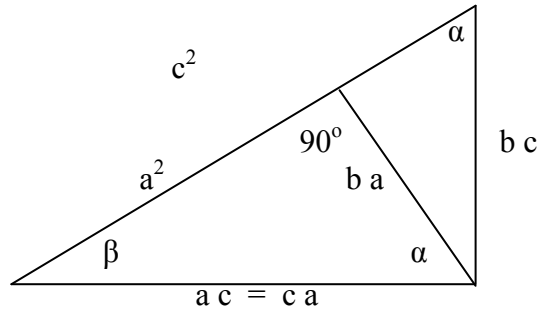
The actual sizes are immaterial but clearly c is the biggest side so that “blow up” should be the biggest and the other two according to the size relations of a and b . So after the student can make a sketch of the three increases at least roughly proportional, we ask if there are any common sides and angles. Of course everybody will see these in all versions. Then we ask the student to place the smaller two versions over the largest c increased one by placing the common sides over each other. There is no possibility of confusion!

Indeed, in the “ a blow up” the only common side is ca being the same as $a c$.

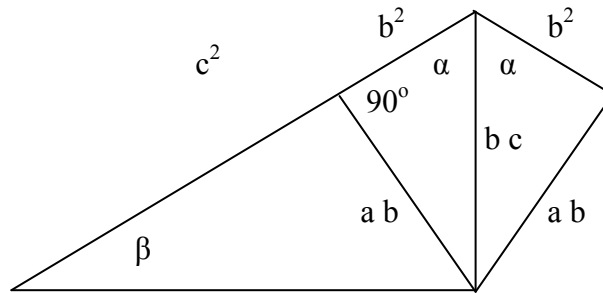
Placing that triangle with its biggest side under the ac side we have it outside but with same β angles on the common sides:



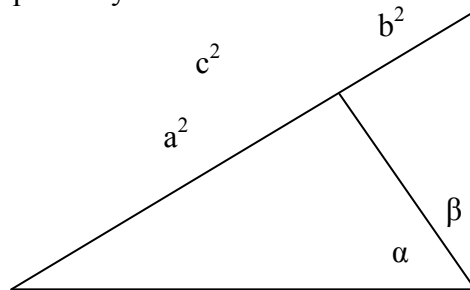
So to cover, we simply have to mirror the lower one into the large:



Similarly, the “b blow up” can be moved next to the b^2 side, then mirrored inside and now the same α angle guarantees that b^2 falls onto c^2 :

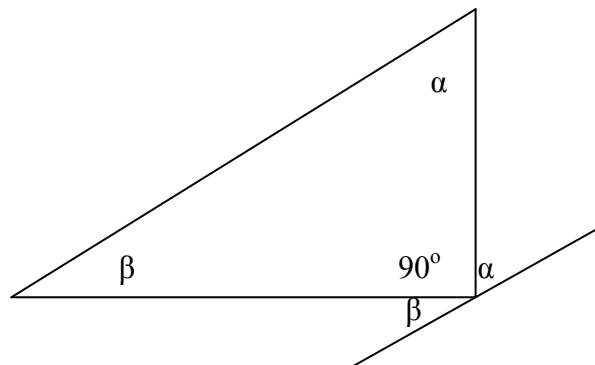


The a^2 and b^2 sides fell onto the c^2 side after the mirrorings in both cases because of the already same β and α angles. But the other b^2 and a^2 sides that are mirrored will not merely be equal but identical. Indeed they both are perpendicular to the c^2 side as marked with the 90° in both pictures. But there is only one perpendicular from the corner across. So in fact the a^2 and b^2 sides not only fall onto the c^2 side but they can't overlap or have a gap, that is they cover perfectly c^2 :



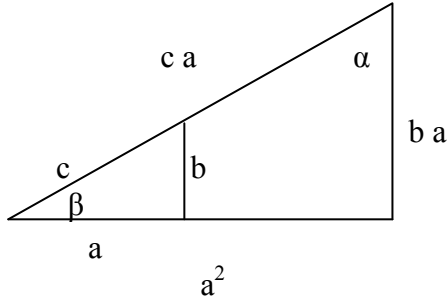
Observe that the argument for the identity of a^2 and b^2 also proved that $\alpha + \beta = 90^\circ$.

Of course a much easier direct proof of this is to draw not a perpendicular rather a parallel to the side across. The two angles then appear under the line so with the 90° already at the corner make a full 180° leaving exactly 90° for α and β :

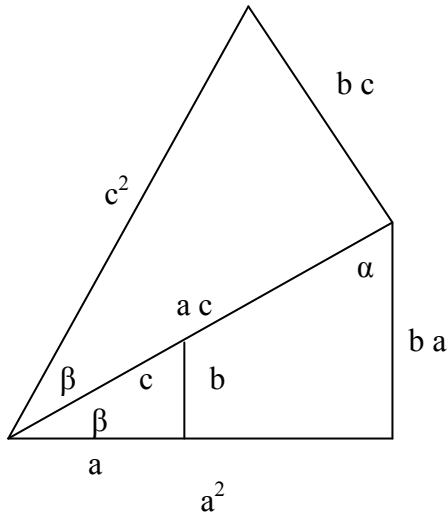


One step closer to a grid oriented proof is the following:

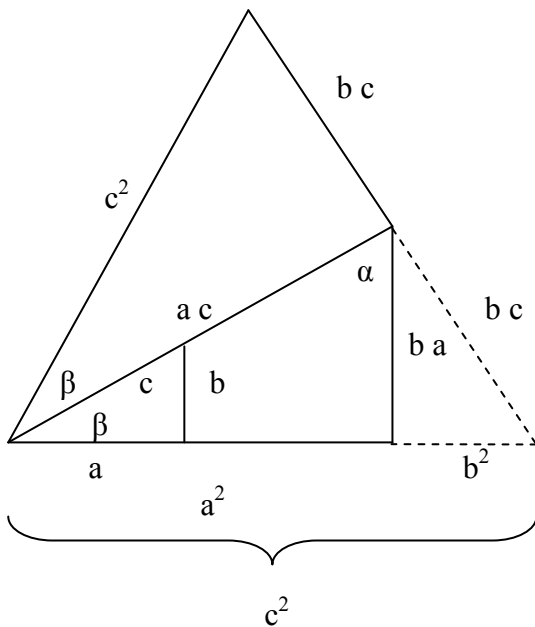
We again increase or blow up the triangle but now first with a staying in the same β angle:



The c blow up is then applied on the c line:



Of course since $c a = a c$ thus this c blow up will be exactly on top and can be mirrored down too and so c^2 will be on the horizontal line covering a^2 :



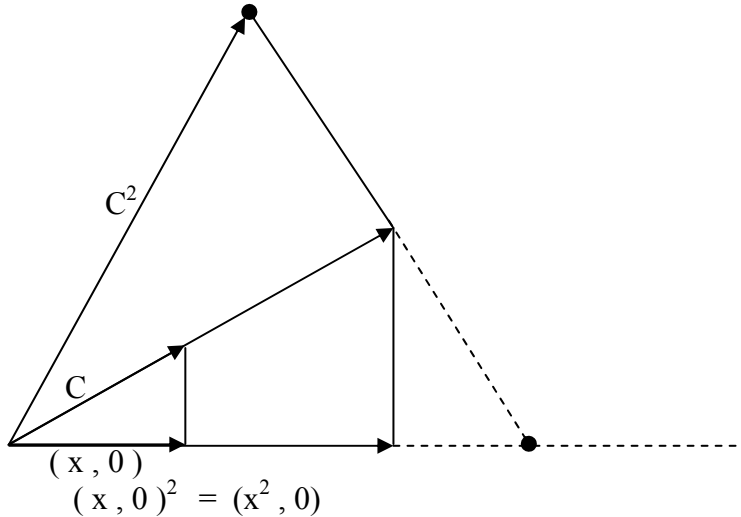
Thus all we have to show is that the leftover in c^2 is b^2 . But the angle across is $90^\circ - \alpha = \beta$, and the sides $b a = a b$, $b c = c b$ are b blow ups of a , c and thus the third is indeed b^2 .

Now we start from connectors:

The $C = (x, y)$ connector with d length is an x, y, d triangle replacing the above a, b, c .

Now the main concern is that x and y are perfect wholes but d is not at all.

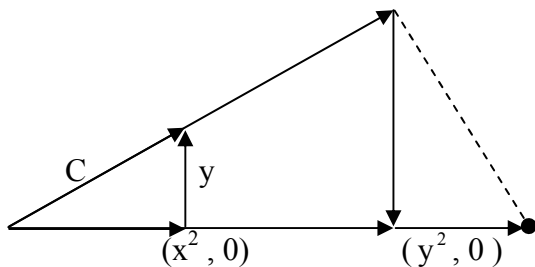
The shocking first claim is that d^2 is always a whole number! To show this, we repeat the previous proof up to the mirroring. In fact now the blow up is proper multiplication with using $(x, 0)$ or C as U new unit, so:



C^2 is a connector with beginning and end on grid points and has the d^2 length but it is still a mystery why this should be a whole length.

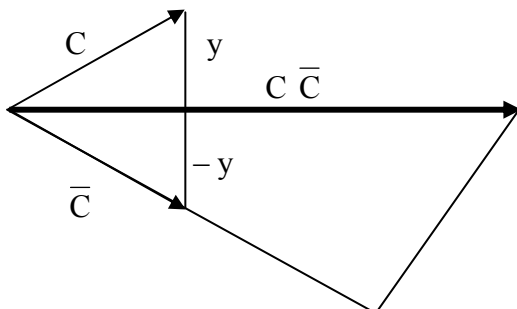
Mirroring C^2 to the line of C , the end of C^2 is mirrored to the end of C . But mirror of grid to an other is again a grid, so it not only falls onto the horizontal line but onto a grid on it! So the self multiplication of a connector, that is measuring it in itself as unit, or sub grid system, always creates a perfect whole length. Thus this d^2 should be regarded as the “value” of a connector, being more important than the d length.

The next step is the sweet surprise that the exact value of a connector is also calculable from the coordinates as $d^2 = x^2 + y^2$. The proof is simply the end of the last, but now again we can be more precise and see that actually the y long but downward, that is $(0, -y)$ connector is used as new unit, creating the y^2 long horizontal and forward, that is $(y^2, 0)$ connector.



So the connector meaning is that $(x^2, 0) + (y^2, 0) = (x^2 + y^2, 0) = (d^2, 0)$.

But this is still not perfectly connector based because this very fact was only shown by the mirroring of C^2 . The really heuristic step is the following: Instead of mirroring C^2 , why don't we mirror $C = (x, y)$ itself? This mirrored $\bar{C} = (x, -y)$ used as new unit would at once put the multiplied C onto the horizontal line and have the same length:



Algebraically: $C \bar{C} = (x, y)(x, -y) = (d^2, 0) = (x^2 + y^2, 0)$

But wait a minute, we already had an algebraic formula for products as:

$$(x, y)(u, v) = (xu - yv, xv + yu).$$

Using $u = x$ and $v = -y$, we get exactly $(x^2 + y^2, 0)$.

So we already proved the Pythagoras theorem but we weren't aware of it!

To sum it up, we gained two important concepts from the Pythagoras Theorem.

The value d^2 and the mirrored $C = (x, -y)$ of any $C = (x, y)$ connector.

The two relate by: $C \bar{C} = (d^2, 0)$.

The seemingly simpler C^2 is actually the more complicated as:

$$C^2 = C C = (x, y)(x, y) = (x^2 - y^2, 2xy).$$

This however has the same length as $C \bar{C}$.

We showed this above geometrically, but it's again verifiable algebraically by having the same values calculated with their mirrored ones. $C \bar{C} = (x^2 + y^2, 0)$ is of course horizontal, so its mirrored is itself but we'll follow the universal method mechanically:

$$C^2 \bar{C}^2 = (x^2 - y^2, 2xy)(x^2 - y^2, -2xy) = ((x^2 - y^2)(x^2 - y^2) + 4x^2y^2, 0)$$

$$C \bar{C} \bar{C} \bar{C} = (x^2 + y^2, 0)(x^2 + y^2, 0) = ((x^2 + y^2)(x^2 + y^2), 0)$$

And indeed the two are the same as easy to verify by calculating the products.

25. Composites and Primes

Composites are non trivial products and primes are not composites and not units.

Among the naturals we had only one trivial product $1 \bullet n$.

Among connectors a product should be trivial if the sub grid system is actually the original.

$U = (1, 0)$ implies this but it is not the only. Indeed, the other three such U are:

$(0, 1)$, $(-1, 0)$, $(0, -1)$. So a composite connector is one that can be expressed as a product not containing these four units as a member.

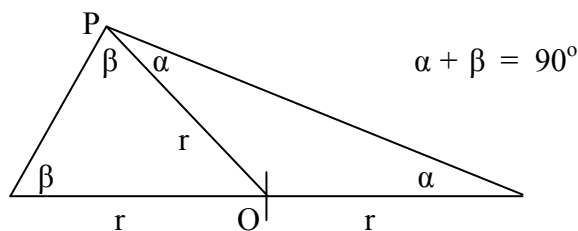
Our heuristic "internal" or "new unit" or "sub-gridding" multiplication started as an opposite to the external stretching or k -multiplication which of course is included in the internal general one as multiplication by $(k, 0)$ horizontal new unit. Strangely though, this original split is still alive in the visual recognition of a composite C connector.

If C is a k -multiple then we can at once see some grid points on C .

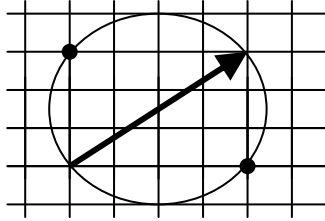
And here the "on" is meant "inside", that is excluding the two trivial end points of C .

If C has no such non trivial grid point on itself, that is, it is a minimal connector, then it can still be composite. But then the new U unit is not on C . Then U measured x times from the beginning of C reaches a P grid point from where a 90° turn and y times U_{90} will get to the end of C . So the C connector looks in 90° from P .

The set of points from which an interval looks in 90° is the Thales circle of the interval, that is the circle having the interval as diameter. This can be proved by connecting the P point with the O middle of the interval, that is with the center of the circle. Indeed the radius of the circle means equal sides in the triangles. Thus the angle at P is half of the total of the angles in the big combined triangle which is 180° :



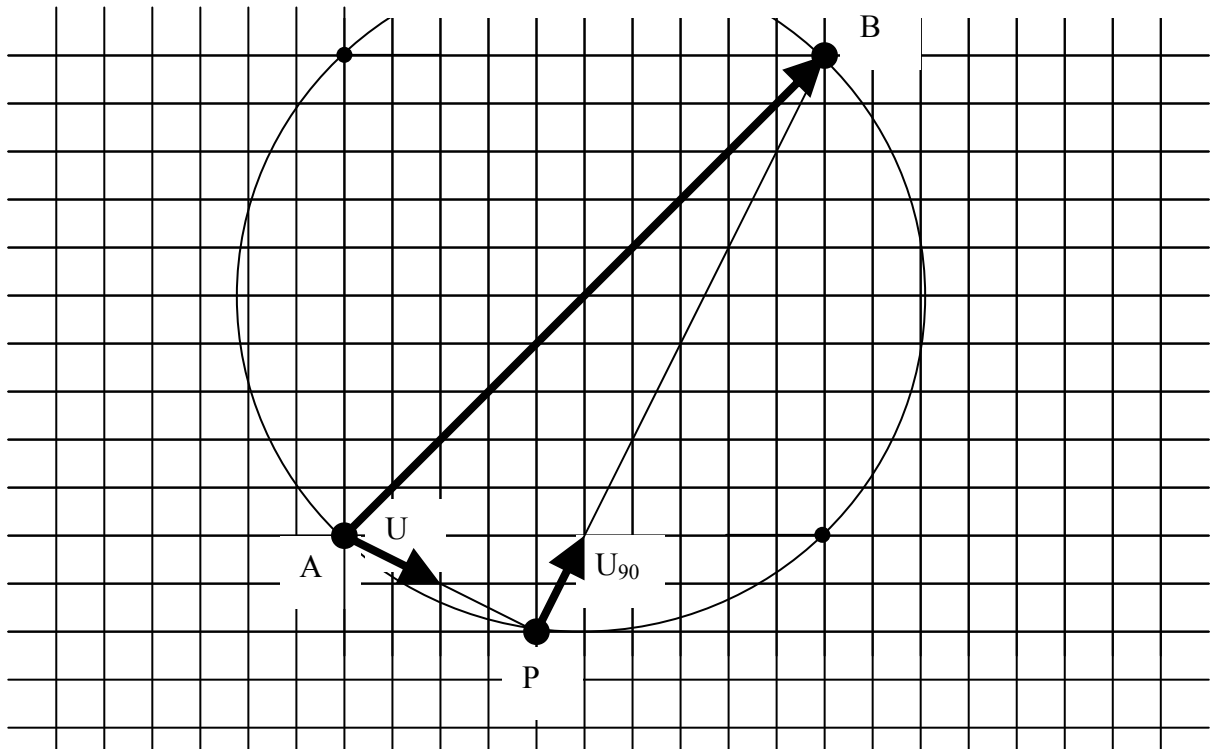
Every non horizontal or vertical connector has two trivial 90° angled grid points on the two sides of the Thales circle, namely the crossings of the horizontal and vertical components:



These are trivial because they don't mean new units only the original grid system.

But any other grid on the Thales circle means an alternative U unit!

Namely, if C goes from the A point to B , that is $C = AB$, then the AP and PB perpendicular connectors will have same minimal connectors on their lines. So the one on AP , starting from A is exactly a new U . Indeed, AP is U or multiple of it and PB is again U_{90} or multiple of it:



For non horizontal or vertical C connectors, these four trivial grids, the two on C itself and the two on its Thales circle, give a very simple visual criteria of compositeness:

A C non horizontal or vertical connector is composite if it has non trivial grid point on itself or on its Thales circle! Or negatively:

C is a prime connector if there are no non trivial grids neither on it nor on its Thales circle.

Perfect algebraic criterias of compositeness or primality were established by Gauss.

The heuristic concept of the d^2 "value" of a C gives a trivial algebraic consequence of compositeness, while the heuristic concept of the \bar{C} symmetrical pair gives a trivial sufficient algebraic condition of compositeness for horizontals.

Indeed, $C U = [d, \alpha] [r, \beta] = [d r, \alpha + \beta]$ so the length of a product is the product of the lengths. We merely used this fact for exchangeability before, because the lengths are not wholes in general only at horizontal and vertical connectors.

But now, in light of the values being wholes, it means more. Indeed, the value of a product is also the product of the values. So a composite connector must have composite value!

For horizontal $C = (x, 0)$ connector, the sufficient condition to be composite is $x = a^2 + b^2$.

Indeed, then $C = (a, b)(a, -b) = (a^2 + b^2, 0)$.

The value consequence of compositeness will be almost perfectly reversible:

A non horizontal or vertical C is composite, exactly if its value $x^2 + y^2$ is composite.

This non horizontal or vertical condition for the reversal is quite plausible. Indeed, every horizontal or vertical connector has composite value except the units, because it's a square.

The sufficient condition of compositeness for horizontals is again only reversible by an extra condition, namely that x is a prime among the naturals. So the reversal simply says that only the square sum p prime numbers loose their primality as $(p, 0)$ connectors.

The non square sum p primes remain prime as $(p, 0)$ too.

Observe that the non horizontal or vertical rules use the value, while the horizontal the length.

It starts to be visible how the connectors could explain the splitting of the natural primes.

But we are still pretty far. We have to continue to explore the connector primes first.

26. Prime factorization, Grid triangles

A prime factorization is easy to obtain! We simply break down the composite more and more till it has to become a product of primes. The tricky part is to prove that this is unique.

No matter how we do the break downs, the end set of primes are the same. The simple fact that proves this is the external atomness of primes. They divide products separately!

Among naturals it meant that if a p prime divides the $m n$ product then p must divide m or n or maybe both. We became more accurate by generalizing this law. We allowed any d divider instead of p and fixed an n that has no common divider with d except 1. Then we could tell that d must divide m . This of course implies the law for $d = p$ primes at once.

The real beauty behind this more general law was that it had an elementary school meaning too.

First we can turn the dividability into a more exact meaning! Indeed, d dividing $m n$ actually means alternative products that is $m n = c d$. To be precise, this should be proved. But we can

go oppositely, back to division again by $\frac{n}{d} = \frac{c}{m}$. So the $\frac{n}{d}$ fraction has an alternative or

variant. The condition that n and d had no common divider simply means that they are simple.

They are simplified completely. And then the claim that d divides m also implies that n divides c , so the variant fraction is merely a multiple version. In elementary school we never make this claim exactly, we merely assume it subconsciously. We feel that equal fractions if simplified, become identical, they are merely expansions of a common simple or one is already the simple.

The heuristic but totally ad hoc idea of the conventional proof is to forget about the n, d condition of being simplified and instead regard all possible equal fractions. Among these, there is a smaller or bigger relationship meaning not their values that is the same of course, rather the nominators or denominators. Indeed they are smaller or bigger simultaneously. But then there has to be a minimal among the equal ones, that is the variants. Then it's easy to see that all variants have to be multiples of this minimal!

But then I showed how connectors allow an almost trivial and yet exact proof too, using the parallelity of the plane. The crucial heuristic idea is that equal fractions always mean that they have identical common expansions. These correspond to multiples of the two fractions as connectors. But multiples are parallel, so both connectors are parallel with the common multiple and thus with each other too. Then of course they can be shifted to a common line and will have the minimal connector of that line inside them.

In spite of this, I also repeated the complicated proof among the connectors and even created a "monstrosity" about rectangles.

Among connector products everything seems to start the same way. The Unique Prime Factorization again boils down to the atomness of P prime connectors, that is that they divide $M N$ products separately. Again, we can go from P prime to any D divider that has no common divider with N except the four units of the grid system. And the claim is that then D will divide M . Again this means that $M N = C D$ so we have alternative products. And again

we can go back to division by $\frac{N}{D} = \frac{C}{M}$.

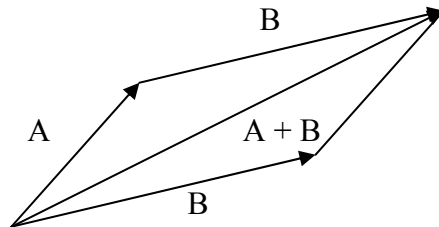
The fractions were an elementary extensions of the naturals to solve equations. But now these connector fractions seem to be artificial. We might hope that just as the fractions found a meaning as connectors and this gave the edge of parallelity to show the unique simplifications and thus the whole atomness of primes too, here again there is a wider picture where everything becomes visible. But it is not the case! The plane is already the wider picture!

Luckily this is still much wider than merely the connectors, so the equal connector fractions do have a perfect visual representation as similar grid triangles. But there is no meta parallelity that would align these fractions or triangles. Or maybe there is but I am too stupid to see it yet.

So all that remains is to apply the ugly indirect arguments of minimalities:

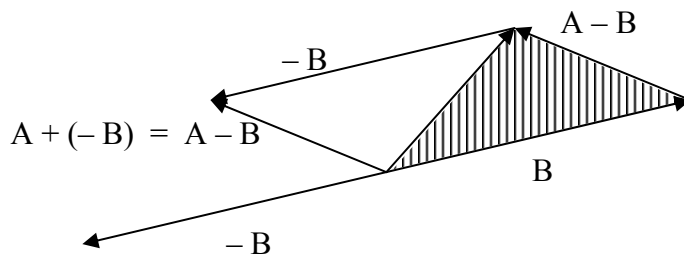
The similar triangle representation of the equal fractions is so beautiful that it makes even this ugly indirect argument beautiful. Plus it contains more surprises too.

We simply imagine the N nominator and D denominator connectors initiated from a single point. This vision is already called among vectors as the “triangle rule” but it is not this crucial new fraction concept rather a visualization of the subtraction or differences. We merely called the addition as continuation of the connectors. Usually it is called the chain rule. The beauty of this is that it applies to arbitrary many members. If two connectors are initiated from one point then only two members can be added visually by the so called parallelogram rule:

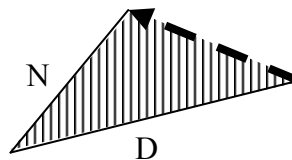


The only reason this is important is that for forces, this is more natural. Like two horses pulling a wagon in angles, this will show the resulting force that moves the wagon.

On the contrary the difference is better seeable from common initiation:



This difference triangle will actually be used in our argument too, but the start is a deeper triangle of the N and D . We can even connect the end points exactly as for the difference but we are not interested in the $N - D$ difference, so we should use a dotted arrow that itself shows the end points of N and D and thus we can omit those:



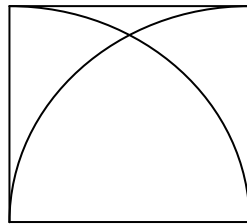
This dotted arrow represents a change from D to N but not in mere difference rather alteration in general that is what angle turning and increase of length occurred. This is exactly what the heuristic concept of multiplication does in forward direction with whole grids. But it is not expectable that N is always a version of D with a new U unit that is $N = D U$. Of course if so, then U is logically $\frac{N}{D}$ as division with perfect connector result. But if not, then still the change from D to N is visible by the angle between them and from the proportion of their length. In fact this is exactly the shape of the triangle between them. So similar triangles indeed should mean equal connector fractions.

It's obvious that among the similar grid triangles there has to be a minimal, say $\frac{N}{D}$.

But now the claim that all equal variants are merely multiples of this minimal can't be as simple as before. Indeed the k -multiples of it are merely external stretchings, staying in the same position. But in a grid system there can be turned versions too. So the heuristic step is to assume that the new internal multiplication can step in and rescue its own development.

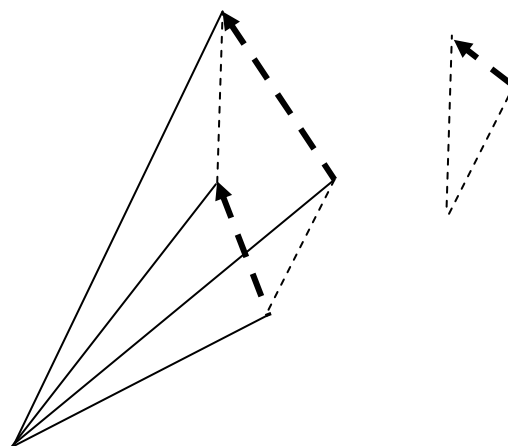
So instead of k -multiples we should regard K -multiples, that is versions obtained by multiplications with K connectors. But this is a huge difference! Indeed the K -multiples don't form an increasing sequence where the exclusion of an intruder could be defied by a simple inbetweenness. But also observe that the real indirect step was merely that an intruder would have to be closer to a multiple than the minimal and so if the differences of equal fractions that is triangles are again such then we do get a contradiction. In a sense the whole indirect argument became better because it really chases the essence of the contradiction.

We might even jump to the conclusion that the K multiplication will be by a sub grid system generated by K . But we are wrong! Instead, we regard two simultaneous sub grid system by the N and D minimal nominators and denominators. So quite oppositely as we thought the K will be measured in these as an arbitrary connector. These will give the multiple versions of the numerators and denominators and thus the multiple triangles too. Of course the two systems are in synchron because a common K must be applied in them. Any intruder, that is not K -multiple triangle would have to be from sides that are not exact connectors in these sub systems. Moving the beginning to an exact grid then the end point would have to be inside a square that has at the four corners exact K -multiple grids. The squares of course have N or D sides in the two subsystems. But a point in a square must be closer to at least two sides than the side:



So indeed the intruder triangle would be closer to a K -multiple than N and D .

Now comes in the old fashioned difference triangle. That shows that differences of similar triangles are similar too:



Thus the intruder would create a contradictory smaller triangle than the minimal:

27. The reversals

The easier is the horizontal one. We saw the trivial general condition:

If x is square sum a^2+b^2 , then $(x, 0)$ is composite namely as $(a, b)(a, -b)$.

The reverse is only true among primes, that is if $(p, 0)$ is a composite then $p = a^2+b^2$.

If any $(n, 0)$ is a composite as $(n, 0) = C U$ then C and U must be symmetrically angled, that is $C = [d, \alpha]$ and $U = [r, -\alpha]$.

For $(p, 0) = C U$ neither C nor U can be k -multiples because then $(k, 0)$ would divide $(p, 0)$ implying that k must divide p contradicting that p is prime as natural.

So C and U must be minimal connectors.

But symmetrically angled lines have same minimals, so C and U must be symmetrical.

So if $C = (a, b)$ then $U = (a, -b)$ and so $(p, 0) = (a, b)(a, -b) = (a^2+b^2, 0)$.

We are finished but an extra fact is that the symmetrical C and U are not merely minimals but actually primes too. Indeed if C were composite then U were too, with symmetrical dividers.

Thus $C U$ would have two symmetrical dividers giving an $(n, 0)$ product dividing $(p, 0)$.

So horizontal and similarly vertical prime long connectors are all either prime connectors or if not then they are the product of two symmetrical primes. The two case merely depends on whether the p prime length is a square sum or not. If not then the prime long connector is prime as connector too.

This settles the splitting of the primes in relation to the bigger picture of connectors but said nothing about the crucial $4k+1$ or $4k-1$ coincidental split. We'll come to that soon.

Checking the square sumness of primes we can see that $2 = 1^2+1^2$ so it is not a prime horizontally or vertically. Instead must be the product of two symmetrical ones. And indeed:

$$(2, 0) = (1, 1)(1, -1)$$

The next prime 3 is not square sum so the 3 long horizontal or vertical connectors are primes.

Next is $5 = 1^2 + 2^2$ so it is again a product of primes as $(5, 0) = (1, 2)(1, -2)$.

A seemingly different prime factorization is $(5, 0) = (2, 1)(2, -1)$.

But observe that: $(2, 1) = (1, -2)(0, 1)$ and $(2, -1) = (1, 2)(0, -1)$.

These can be seen by the very meanings of products or by the coordinate formula too.

So these are merely unit variations because $(0, 1)$ and $(0, -1)$ are units.

By the way the multiplication with them meant 90° and -90° turns.

You can continue and play with these prime verifications but soon we'll come to a much easier fast calculating method for connector products.

The crucial non horizontal or vertical reversal is the most beautiful result of connectors.

Remember, the trivial rule was that composite C must have composite d^2 value.

So the reversal is that if d^2 is composite then C is too.

First of all for k -multiple C 's it's trivial, so enough to show the claim for minimal C 's.

We will be much more concrete and actually tell how a $d^2 = p_1 p_2 \dots p_m$ prime factorization of d^2 leads to a prime factorization of the minimal C itself.

First of all, for any of these p_i prime factors of d^2 , the $(p_i, 0)$ horizontal connector can not be prime. Indeed, p_i dividing d^2 also means $(p_i, 0)$ dividing $(d^2, 0) = C \bar{C}$.

So $(p_i, 0)$ being prime would require to divide C or \bar{C} . But these being minimal, can not be divided by $(p_i, 0)$ because that would mean to be p_i -multiple too

So every $(p_i, 0)$ is actually a $P_i \bar{P}_i$ product of two primes as we established above.

So then $C \bar{C} = P_1 \bar{P}_1 P_2 \bar{P}_2 \dots P_m \bar{P}_m$. But here again, every P_i dividing $C \bar{C}$ must divide C or \bar{C} maybe both. Of course if it divides say C then the mirrored \bar{P}_i must divide \bar{C} . So indeed we get a simultaneous prime factorization of C and \bar{C} .

28. The simple square sums

Any $x^2 + y^2$ square sum means that $(x^2 + y^2, 0) = (x, y)(x, -y) = C \bar{C} = (d^2, 0)$.

Furthermore $x^2 + y^2$ being simple means exactly that x, y are simple that is without common divider, that is C and \bar{C} are minimal. So the proof of the last reversal was a proof about simple square sums too. Namely we proved that every p_i prime factor of $d^2 = x^2 + y^2$ is a non prime as horizontal, thus it is $P_i \bar{P}_i$ and so is an $a^2 + b^2$ itself. Of course it has to be simple too because it is a prime number. Thus we obtained the most important earlier claim that the prime factors of simple square sums are the same. By the square remainders we also established that every $4k + 1$ prime number is factor of some $r^2 + 1$. This of course is a square sum because $1 = 1^2$. And simple because 1 has no other divider than 1.

So the $4k + 1$ primes are square sums.

We did not tell anything more about the p_i prime factors of $x^2 + y^2$ beside that they have to be $a^2 + b^2$ themselves. So the natural question is whether any such square sum primes multiplied together will give a simple square sum in reverse. If you remember, we already revealed that this is true only almost, namely the 2 prime factor can only be used once. Why?

A product of symmetrical $P_i \bar{P}_i$ primes obviously will create a symmetrical total and thus a square sum. So the only problem could be that the simpleness is ruined. Remember that simpleness was defined as having such square sum form, not having only such. Having any repeated p prime factor means that p^2 can be taken out and if the left number is square sum as $u^2 + v^2$ then we at once have a non simple form as $(pu)^2 + (pv)^2$. But this doesn't mean that an other simple can't be too. This is always the case with the non 2 prime factors. The smallest 5 for example repeated plus 2 gives $50 = 5^2(1^2 + 1^2) = 5^2 + 5^2$ is a non simple form. But of course we have the simple $1^2 + 7^2$ too. If however a number contains more than two 2 factors, that is dividable by 4, then we can never find a simple form. Why again?

In the proof of the previous section we didn't separate the $P_i \bar{P}_i$ pairs by our will because the already existing C and \bar{C} minimal told this. Only if a P_i was dividing both C and \bar{C} could we have a choice and we went in order, so claimed it as a C factor.

Now we have to separate the arbitrarily chosen $P_i \bar{P}_i$ pairs by a method that guarantees minimal that is simple final C and \bar{C} . So the C and \bar{C} are like empty baskets that we fill up by our method. It's also obvious how too to get simple product, namely we have to avoid mirrored versions in a basket. Only these can create a natural divider of C or \bar{C} .

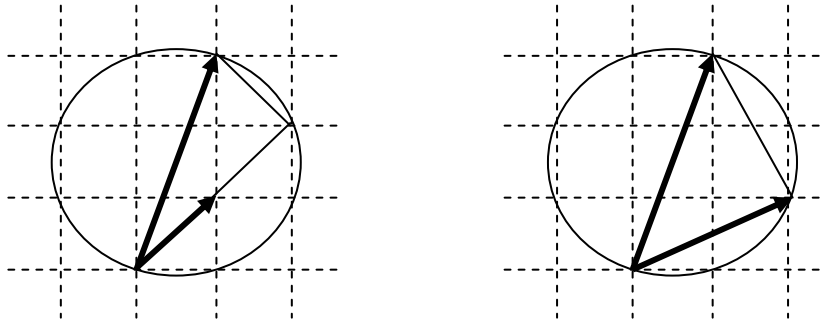
The method is then this:

When a $P_i \bar{P}_i$ pair appears the first time then we can pick freely, so simply pick the first into C . If however it repeats, we must put the same member of $P_i \bar{P}_i$ into C as we did before.

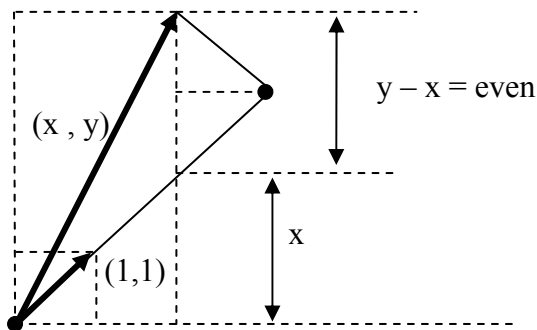
An other way of doing all this is to order the repeating $P_i \bar{P}_i$ pairs inside so that they are not merely equal products but identical too and then just always pick the first members into C .

But remember, "same" or "identical" among the connectors as product members now must include the unit variants, that is the 90° , -90° and 180° turns. If this interferes with the $P_i \bar{P}_i$ pairs then we might not be able to choose. These pairs are mirrored ones. So can a mirrored be also the 90° or -90° or 180° turned? Unfortunately yes, for 45° it is the same. So is there prime connector with this angle? Yes the simplest prime is $(1, 1) = [\sqrt{2}, 45^\circ]$. Of course it is the same in products as its unit variants: $(-1, 1)$, $(1, -1)$, $(-1, -1)$. Indeed if a mirrored pair of these appear, then picking the first or second is the same, so we cant pick again avoiding a mirrored one getting in the same basket. That's why $(1, 1)(1, -1) = 2$ can be allowed as prime factor of $x^2 + y^2$ only once to keep simplicity.

This already shows that the $(1, 1)$ prime connector corresponds to the 2 among naturals. The analogy goes much further! We can define parity among the connectors namely as being dividable by $(1, 1)$ or not. But most amazingly this can be seen from the coordinates at once too as having the same normal parity or not. The simplest minimal connector that is not prime is $(1, 3)$ and it is “even” because the coordinates have same parity namely both odds. Thus it has to be dividable by $(1, 1)$ and indeed: $(1, 3) = (2, 1)(1, 1)$ or $(1, 1)(2, 1)$:



To prove the general claim:



29. The fast connector arithmetic

The earlier mentioned fact that every C connector is actually the sum of its components, that is $C = X + Y$ didn't seem like a more practical form of C than (x, y) . Amazingly if we over complicate thing even further, then we get something really simple in the end!

The X and Y can be themselves obtained from the units and the coordinates so:

$C = X + Y = x(1, 0) + y(0, 1)$. Writing all connectors in this form we can indeed calculate with them merely obeying the old rules of additions and multiplications. Only when multiplications of these two units appear among themselves must we go outside the old rules.

But in fact from the four possible product involving $(1, 0)$ and $(0, 1)$ only one is really outside namely $(0, 1)(0, 1) = (-1, 0)$. Indeed in the other three cases where $(1, 0)$ is involved we can simply omit that. The addition of $(-1, 0)$ or its multiples is of course the same as the subtraction of $(1, 0)$ so to put it simply: $(-1, 0) = -(1, 0)$.

So we can remain completely inside using only $(1, 0)$ and $(0, 1)$.

A natural idea would be then to introduce some abbreviations for these two, but since $(1, 0)$ in multiplications is omitted anyway, thus we only need one symbol for $(0, 1)$ if we agree that all numbers appearing by themselves include the $(1, 0)$. In short: $(1, 0) = 1$

This of course also means that $(0, 1)(0, 1) = (-1, 0) = -1$

Using i for $(0, 1)$ then we have: $C = X + Y = x(1, 0) + y(0, 1) = x + yi$.

Calculations become mere algebraic continuations of the old rules except for the new: $ii = -1$.

The old rule of multiplications for example becomes also derivable algebraically as:

$$(x, y)(u, v) = (x + yi)(u + vi) = xu + xvi + yiu + yivi = xu + yvii + xvi + yui = (xu - yv) + (xv + yu)i = (xu - yv, xv + yu)$$

But for concrete numbers, the method is even more amazingly practical. The factorization of the above mentioned $(1, 3)$ is for example: $(2 + i)(1 + i) = 2 + 2i + i - 1 = 1 + 3i$.

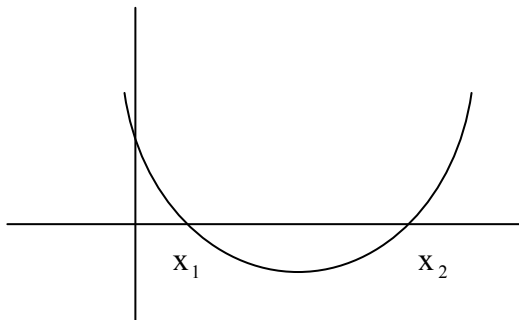
30. The problem of the Fundamental Theorem of Algebra

These $(x + yi)(u + vi)$ multiplications were known way before Gauss.

It had nothing to do with primes and square sums, rather with solving equations.

The second order equations are taught in every high school, which is understandable because many word problems lead to such $ax^2 + bx + c = 0$.

Since the coordinate system is also part of every high school math curriculum, it's logical to show that the picture of a $y = ax^2 + bx + c$ is a parabola.



The position and fatness or narrowness of the parabola depend on a , b , c , and it's easy to see that if a is negative, then the parabola is upside down. It's harder to tell whether a crossing with the x axis is happening or not. But this is vital, because it means the $y = 0$ points, that is exactly the solutions of $ax^2 + bx + c = 0$. On the other hand, a third obligatory subject in high school, the method of finding these solutions is totally unrelated to the parabola picture.

Back in the sixties when I went to high school, in most math classes, they merely gave the final solution formula:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Kids had to learn this and use it for the examples.

Then in a more "modern" method, the first step is that a is not really important for the roots. Indeed, we can divide with it because its never 0, otherwise it wouldn't be a second order equation. So, $x^2 + \frac{b}{a}x + \frac{c}{a} = 0$ should be regarded. This is not only simpler, because it has only two data, $\frac{b}{a}$ and $\frac{c}{a}$, but these are directly related to the solutions.

This is easy to obtain from assuming two x_1 , x_2 solutions and then:

$$\underbrace{\left(x_2^2 + \frac{b}{a}x_2 + \frac{c}{a}\right)}_0 - \underbrace{\left(x_1^2 + \frac{b}{a}x_1 + \frac{c}{a}\right)}_0 = x_2^2 - x_1^2 + \frac{b}{a}(x_2 - x_1) = 0$$

Now observe that $(x_2 - x_1)(x_1 + x_2) = x_2x_1 + x_2^2 - x_1^2 - x_1x_2 = x_2^2 - x_1^2$.

So, $(x_2 - x_1)(x_1 + x_2) + \frac{b}{a}(x_2 - x_1) = 0$.

Dividing it with $x_2 - x_1$ and then subtracting $x_1 + x_2$ from both sides we get:

$\frac{b}{a} = -(x_1 + x_2)$ Putting this back in any of the used x_2 or x_1 equations:

$x_2^2 - (x_1 + x_2)x_2 + \frac{c}{a} = x_2^2 - x_1x_2 - x_2^2 + \frac{c}{a} = 0$. Thus, $\frac{c}{a} = x_1x_2$

First of all, in some schools, they don't explain all this, merely tell that:

$-\frac{b}{a} = x_1 + x_2$ and $\frac{c}{a} = x_1x_2$, that is the sum and product of the two roots are directly the two data of the simplified equation. Then, give idiotically simple examples where from the sum and the product, the students are supposed to guess the roots.

For example in $x^2 - 7x + 12 = 0$ $x_1 + x_2 = 7$, $x_1 x_2 = 12$ so $x_{1,2} = 3, 4$.

Of course, for any reasonably hard example, such guessing is hopeless.

The education system goes through these stupidity cycles. In the sixties, it was stupid to force the solution formula, fair enough. But then, they at least solved word problems and used it. Then the idiotic education egg-heads wanted to be "investigative" and gave a new method that is useless in practice. So text books are filled with hundreds of pre-digested equations that can be guessed, but of course not one real word problem leads to them.

This $x_1 + x_2 = -\frac{b}{a}$, $x_1 x_2 = \frac{c}{a}$ rule, actually could be used positively in two ways.

Firstly, it is a two variable equation system for x_1, x_2 and so would be a practice for that. Namely, it would be a very useful unsuccessful attempt. Indeed, in spite of the simplicity of it, if we try to solve it, that is express one say x_2 from the first equation

$x_1 + x_2 = -\frac{b}{a}$ then $x_2 = -\frac{b}{a} - x_1$ but writing this into the second, we get:

$x_1 x_2 = x_1(-\frac{b}{a} - x_1) = -\frac{b}{a} x_1 - x_1^2 = \frac{c}{a}$ so, $x_1^2 + \frac{b}{a} x_1 + \frac{c}{a} = 0$ and thus, we literally got back to "square" one.

And yet amazingly, the subjective simplicity of the $x_1 + x_2 = -\frac{b}{a}$, $x_1 x_2 = \frac{c}{a}$ equations,

does pay off. All we have to do is go back a few thousand years.

Indeed, the Babylonians solved this problem. Lets be a bit contemplative. We are given the sum and product of two unknown quantities: $x_1 + x_2 = s$, $x_1 x_2 = p$

Now what else is such simple thing? Obviously, the difference: $x_2 - x_1 = d$

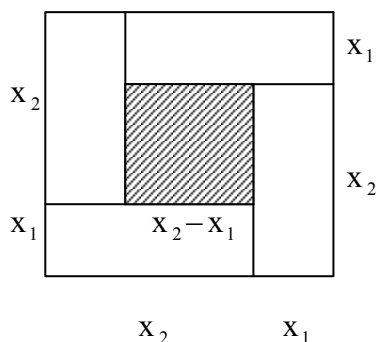
What's more, if we knew this difference, we were finished at once, because the sum and difference give easily any two quantities:

$$x_1 = \frac{s-d}{2} = \frac{(x_1+x_2) - (x_2-x_1)}{2} = \frac{2x_1}{2} = x_1$$

$$x_2 = \frac{s+d}{2} = \frac{(x_1+x_2) + (x_2-x_1)}{2} = \frac{2x_2}{2} = x_2$$

But how could we get d somehow?

The Babylonians came up with a Hindu proof very closely resembling the Pythagoras one:



$$(x_2 - x_1)^2 = (x_1 + x_2)^2 - 4x_1 x_2$$

$$d^2 = s^2 - 4p$$

$$d = \sqrt{s^2 - 4p}$$

So, from s and p its quite easy to calculate d . In fact, using $s = -\frac{b}{a}$ and $p = \frac{c}{a}$ and then d and s to get x_1 and x_2 , we easily obtain the formula for the second order equation:

$$d = \sqrt{s^2 - 4p} = \sqrt{\frac{b^2}{a^2} - 4\frac{c}{a}} = \sqrt{\frac{b^2 - 4ac}{a^2}} = \frac{\sqrt{b^2 - 4ac}}{a}$$

So:

$$x_1 = \frac{s-d}{2} = \frac{-\frac{b}{a} - \frac{\sqrt{b^2 - 4ac}}{a}}{2} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

$$x_2 = \frac{s+d}{2} = \frac{-\frac{b}{a} + \frac{\sqrt{b^2 - 4ac}}{a}}{2} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

Of course, the Babylonian trick today has a mere algebraic solution too:

$$(x_2 - x_1)^2 = x_2^2 + x_1^2 - 2x_2x_1, \quad (x_1 + x_2)^2 = x_1^2 + x_2^2 + 2x_2x_1$$

$$\text{So indeed, } (x_2 - x_1)^2 = (x_1 + x_2)^2 - 4x_1x_2$$

Back in the sixties when the second order formula was really used for word problems, a very frequent question was whether the two solutions are both real solutions of the problem or not.

Sometimes for example, one became positive, the other negative for a distance problem that had to be positive. So it was just assumed that a meaningless solution can come about too. Even more interesting is if another problem leads to a second order equation, so a , b , c depend on some other quantities. Then one could use the solution formula, but in it $\sqrt{b^2 - 4ac}$ must be assumed to become meaningful, that is $b^2 - 4ac \geq 0$. By the way, this is also the condition of the parabola crossing the x axis. So the dilemma is this: Is this condition merely a perfect condition of reality conditions too, or the solution formula carries some meaning even when the square root can't exist? Is there a reality beyond the realities that we directly aim for?

Here in the second order case, it's not so critical, because $b^2 - 4ac < 0$ can make us stop right away. But more complicated formulas did come up with negatives under the square root, where those cancelled each other and at the end, meaningful results could come out, even from such impossible cases. This desperate cluelessness, about the negatives under square root went on for hundreds of years. Of course, it was obvious, that the "root" of this square root problem is merely $\sqrt{-1}$. Indeed, for example $\sqrt{-5} = \sqrt{5} \sqrt{-1}$.

The only exact fact we know about thus "imaginary" number $\sqrt{-1}$ is that its square is -1 .

Or to be even more concrete the fact is merely: $\sqrt{-1} \sqrt{-1} = -1$.

So the i short abbreviation for $\sqrt{-1}$ came from imaginary and its rule was: $ii = -1$.

But $\sqrt{-1}$ or i can be involved in all kinds of complicated expressions and it wasn't at once narrowed down to the $u \pm vi$ dual combinations.

In fact, the second order formulas we used is also hiding it when $b^2 - 4ac < 0$:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b}{2a} \pm \frac{\sqrt{(4ac - b^2)}}{2a} i$$

As we see, the roots became symmetrical pairs just as the connectors that have a natural product. But the real hint towards the $u \pm vi$ forms as basic units, came with the recognition that this law is universal to any order algebraic equations.

A fifth order is for example, $3x^5 + 2x^4 - x^3 + 3x^2 + 5 = 0$.

In general, $c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0 = 0$.

The c -s are called the coefficients.

The most educational is first to calculate just the third power of $u + vi$ and $u - vi$.

$$(u + vi)^3 = (u + vi)(u + vi)(u + vi) = u^3 + 3u^2vi + 3u(vi)^2 + (vi)^3$$

Indeed, there are three ways u, u, vi can be picked, and also three ways u, vi, vi .

Now at $(u - vi)^3$ only the single $(-vi)$ and the $u, u, -vi$ give negative products, so

$$(u - vi)^3 = (u - vi)(u - vi)(u - vi) = u^3 - 3u^2vi + 3u(vi)^2 - (vi)^3.$$

Using $i^2 = -1$ and $i^3 = i^2 i = -1 i = -i$.

$$(u + vi)^3 = (u^3 - 3u^2v^2) + (3u^2v - v^3)i \quad \text{and} \quad (u - vi)^3 = (u^3 - 3u^2v^2) - (3u^2v - v^3)i$$

So the mirrored $u + vi, u - vi$ pairs give mirrored values.

This goes similarly for any powers, their multiples and sums too. So the end result is that:

Algebraic $A(x) = c_n x^n + \dots + c_0$ gives mirrored values for mirrored $u + vi, u - vi$.

Now if $u + vi$ is a formal root of $A(x)$, that is:

$$A(u + vi) = (\quad) + (\quad)i = 0 \quad \text{then both} \quad (\quad) \quad \text{have to be } 0. \quad \text{Thus:}$$

$$A(u - vi) = (\quad) - (\quad)i = 0 \quad \text{too.}$$

Beside this grand rule of symmetry, there was an equally grand formal process known for algebraic expressions, namely the exact copy of division among numbers.

The trick is just to regard the first coefficients as we proceed, member by member:

$$(3x^5 + 2x^4 - x^3 + 3x^2 + 5) : (2x^2 + x - 1) = \frac{3}{2}x^3 + \frac{1}{4}x^2 + \frac{1}{8}x + \frac{25}{16}$$

$$3x^5 + \frac{3}{2}x^4 - \frac{3}{2}x^3$$

$$0 \quad \frac{1}{2}x^4 + \frac{1}{2}x^3 + 3x^2$$

$$\frac{1}{2}x^4 + \frac{1}{4}x^3 - \frac{1}{4}x^2$$

$$0 \quad \frac{1}{4}x^3 + \frac{13}{4}x^2 + 0x$$

$$\frac{1}{4}x^3 + \frac{1}{8}x^2 - \frac{1}{8}x$$

$$0 \quad \frac{25}{8}x^2 + \frac{1}{8}x + 5$$

$$\frac{25}{8}x^2 + \frac{25}{16}x - \frac{25}{16}$$

$$0 \quad -\frac{23}{16}x + \frac{105}{16}$$

This last is the remainder, so in fact:

$$3x^5 + 2x^4 - x^3 + 3x^2 + 5 = (2x^2 + x - 1)\left(\frac{3}{2}x^3 + \frac{1}{4}x^2 + \frac{1}{8}x + \frac{25}{16}\right) + \left(-\frac{23}{16}x + \frac{105}{16}\right)$$

In general, if the $A(x)$ algebraic expression is divided by a lower order $D(x)$ then we get a $B(x)$ result and $R(x)$ remainder that is: $A(x) = D(x)B(x) + R(x)$

Especially interesting is, if $D(x) = (x - a)$ simple first order expression.

Then, $A(x) = (x - a)B(x) + r$. So the remainder is a mere number.

But most importantly, if a was a root of $A(x)$ that is $A(a) = 0$, then:

$A(a) = (a - a)B(a) + r = 0B(a) + r = r = 0$. So in fact, $A(x) = (x - a)B(x)$.

So dividing with an $x - a$ with a being a root, there won't be a remainder.

In short, $A(x)$ becomes a product. Then, if $B(x)$ has some b root again, then

$B(x) = (x - b)C(x)$.

Of course, if b is root of $B(x)$ then it is root of $A(x)$ too. So continuing this till the end,

we get: $A(x) = (x - a)(x - b)(x - c) \dots c_n$.

All these a, b, c, \dots are roots of $A(x)$ and the last c_n division result is actually the original co-efficient of A , because this was inherited to $B(x), C(x), \dots$

This at once proves that an n -th order $A(x)$ can have maximum n roots. Of course, it can have easily less for two reasons. Firstly, B can have the same a root as A had, so b can be a again, and so on. Secondly, there was no reason to assume the existence of already a , then b and so on. We can be stuck at any point. Of course, the reverse is much more obvious.

If $A(x)$ is such product, $(x - a)(x - b) \dots$ then these a, b, \dots are all roots, simply because one of $x - a$ or $x - b$ or \dots would give 0 values.

This "root" product form was too beautiful not to be true in general, so it was expected that even if there are no real a, b, \dots roots, at least formal $u + vi$ ones must exist. So:

$A(x) = (x - a)(x - b) \dots (x - (u_1 + v_1 i))(x - (u_2 + v_2 i)) \dots c_n$

Then with the earlier result that for every $u + vi$ formal root, $u - vi$ is root too, we can order them so that $u_2 + v_2 i$ is actually $u_1 - v_1 i$. Similarly, putting all of them in pairs:

$A(x) = (x - a)(x - b) \dots (x - (u_1 + v_1 i))(x - (u_1 - v_1 i)) \dots c_n$

But observe that: $(x - (u + vi))(x - (u - vi)) = x^2 - 2ux + u^2 + v^2$.

So thus, every $A(x)$ would become a product of the real root factors and second order factors containing the artificial root pairs. Some doubted this perfection.

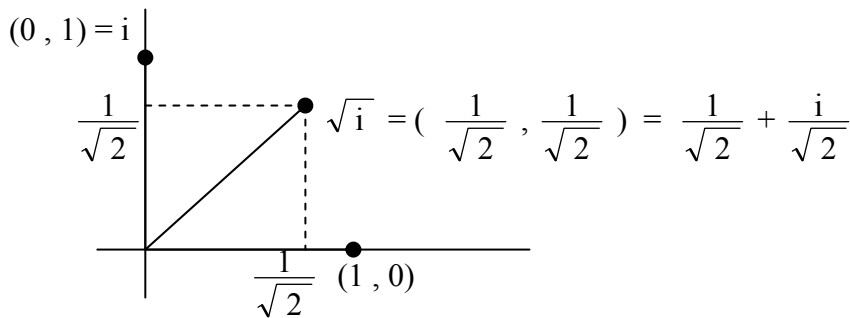
For example, Leibniz claimed that $A(x) = x^4 + 1$ already can't be split.

Obviously, we can't have real roots, because $x^4 = -1$ is just as impossible as $x^2 = -1$.

In fact, x would be $= \sqrt[4]{-1} = \sqrt{\sqrt{-1}} = \sqrt{i}$.

So the thing Leibniz doubted is whether the square root of i is itself an i combination.

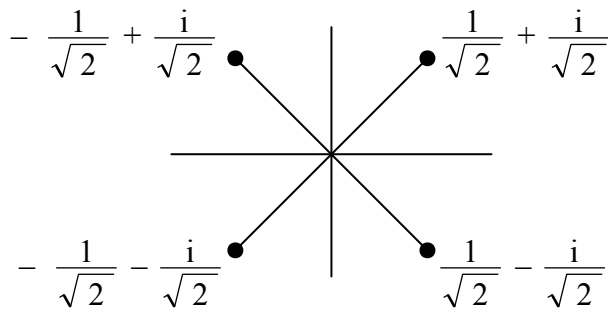
With our modern view it is obviously true:



And indeed, $(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}})^2 = \frac{1}{2} + \frac{i^2}{2} + 2 \frac{1}{\sqrt{2}} \frac{i}{\sqrt{2}} = \frac{1}{2} - \frac{1}{2} + i = i$

By the way, clearly the negative of this must have the same square, that is i .

These two values of \sqrt{i} only give two solutions for $\sqrt[4]{-1}$, but we have the basic rule too, that mirrored ones are roots for real. So in fact, the four fourth root of -1 are:



So $x^4 + 1 =$

$$\underbrace{\left(x - \left(\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)\right) \left(x - \left(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right)\right)}_{(x^2 + \sqrt{2}x + 1)} \underbrace{\left(x - \left(-\frac{1}{\sqrt{2}} + \frac{i}{\sqrt{2}}\right)\right) \left(x - \left(-\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right)\right)}_{(x^2 - \sqrt{2}x + 1)}$$

Euler observed this, thus correcting Leibniz and putting the faith back firmly, that all $A(x)$ can be factorized. Unfortunately, he couldn't prove the crucial existence of a single formal root for all $A(x)$. Euler was especially obsessed with this total factorization, because he realized that it at once proves the generalization of the rule we used at the second order formula for the sum and product of roots:

For arbitrary $A(x)$, again we have to divide with the c_n highest co-efficient, that is avoid it and so if $A(x) = x^n + c_{n-1}x^{n-1} + \dots + c_1x + c_0$ then, $-c_{n-1}$ is the sum of the roots, but now $\pm c_0$ is the product of them depending on n being even or odd.

Indeed, using just a, b, c, \dots, r as roots regardless real or phony, we have:

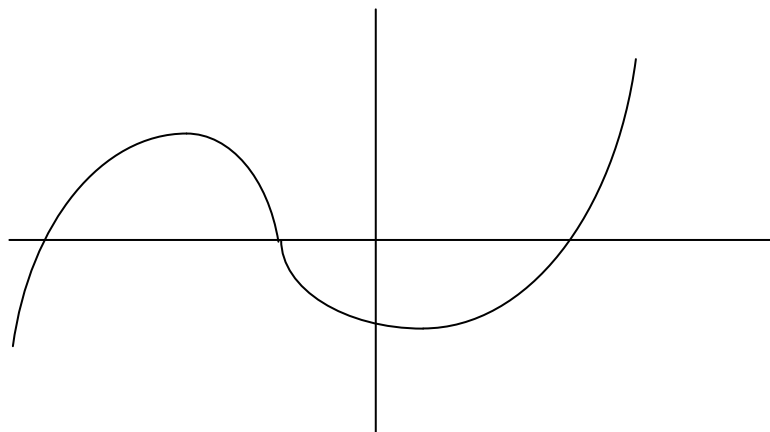
$A(x) = (x - a)(x - b) \dots (x - r)$ Here using all x we get x^n . Using $n - 1$ many x and one root, we get all of them with minus, so $-x^{n-1}(a + b + \dots + r)$. Finally, using all the minus roots, we get their product, but the sign is negative if we use odd many.

The worst thing about the missing evidence for existence of real or artificial roots was that for odd order $A(x)$ it was already trivial that real root exists. For a second, we might say it's okay, because then at least the previous rule for example, is still true for odd order $A(x)$. Not!

Lets remember that the factorization was done step by step, always assuming a root. So we have to go through even ones, even if the original is odd.

At any rate, lets see why a real root is obvious for odd n .

Simply because these $A(x)$ have arbitrary big negative and positive values:



Indeed, let $A(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0$ and c_n be positive.

If x goes towards $-\infty$ then $c_n x^n$ dominates all the other members as a huge negative value, since n is odd. When x goes towards $+\infty$ then of course, the value becomes $+\infty$.

If c_n is negative, then the picture is opposite, going from $+\infty$ to $-\infty$.

In either case, the function has to cross the x axis.

With even n , the $c_n x^n$ is again dominating, but it becomes the same infinity, when x approaches either $-\infty$ or $+\infty$. So, a crossing is not guaranteed.

How can such stupid little difference cause such a deep problem?

No wonder, all mathematicians went nuts about this whole affair.

31. Proving the Fundamental Theory of Algebra

Even when finally Gauss as a young man saw the light in the tunnel, namely that the $u + v i$ combinations are actually the points of a plane as new numbers, he still kept the old distinction of real and non real roots. He didn't transform the whole question into the plane.

That would have been going from $A(x) = c_n x^n + c_{n-1} x^{n-1} + \dots + c_1 x + c_0$ to

$$\mathcal{A}(P) = C_n \circ P^n + C_{n-1} \circ P^{n-1} + \dots + C_1 \circ P + C_0 .$$

This is perfectly meaningful. $P^n = P \circ P \circ \dots \circ P$, that is using P as unit and application n times. This of course means multiplying P 's angle by n and the n -th power of its length.

The C_n coefficients are now points themselves and their application is clear.

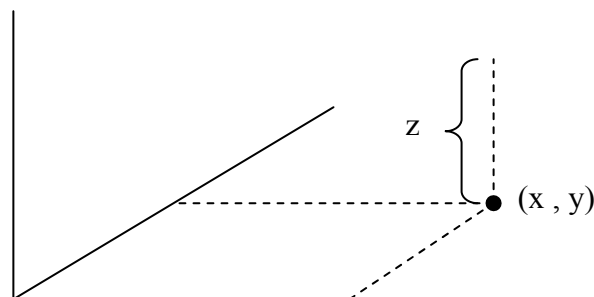
Adding all these turnings together, we get the $\mathcal{A}(P)$ value.

The odd - even distinction of real algebraic expressions or parabolas disappears, but the major fact that $C_n \circ P^n$ dominates the smaller ones remains. So for example, even if C_{n-1} is much bigger than C_n , the $C_n \circ P^n$ will be bigger than $C_{n-1} \circ P^{n-1}$ if P is big enough.

Here of course, "big" is meant by the point's distance from the origin.

While earlier the plane was a subjective tool where mathematicians drew the graph of the real $A(x)$ as $y = A(x)$, now the x, y plane is itself the domain. This weird change of convenience was a strong reason not to realize the transition. Today as well, the Descartes system as a silver platter is so strong in high school and every day life as the field of graphs and charts, that it covers up the realities of functions. In high school, the reality that $y = x^2$ as a parabola is an illusion, because we are dealing with a function that orders numbers to numbers, that is actually in one single line, is never even mentioned.

Now if we use the plane as a domain, then the $Q = \mathcal{A}(P)$ creation of Q points from P , couldn't be graphed in a similar way as the Descartes system, even in space. Indeed, three dimension can only order single z values to (x, y) points.



To "see" $Q = \mathcal{A}(P)$, we would need four dimensions. But this is the stupidest complaint!

Giving up the graph convenience is a step towards truth!

A truth forced upon us by the three dimensionality of our space.

I'm not saying that we should give up Descartes charts on economic news bulletins, but I do claim that when the domain is the plane, then it is didactically ripe to go back to reality and visualize the $Q = \mathcal{A}(P)$ orderings in the single plane.

This will involve time as a new false illusion in a strange multi-levelled way.

The continuity of $\mathcal{A}(P)$ means first of all that moving P to new points, the Q value moves too. So here time is only potential. The usual “spaced” version of this vision is that the set of P values, the domain is the infinite plane as a rubber sheet and the Q values are a bent, wrinkled, stretched version. So we don’t visualize how the points of the rubber sheet moved, only the initial flat and final transformed version is important.

This already helps a lot to see the major problem.

If for example, we have a finite round table and the rubber sheet is glued on the edge, then no matter how we deform it on the table, it will always cover. If the edge is not glued, then of course, we can fold it over and leave half of the table uncovered. Simplest way to do this is by folding the edge of the rubber sheet on the other half. Then, we can stretch the half doubled edge of the rubber sheet to cover the whole edge of the table, but still leaving some uncovered part on the table. If however, we are only allowed to move the edge of the rubber, on the edge of the table around, then again, the covering must remain. Now if the “history” of the rubber sheet doesn’t count, then what’s the difference between the folding and re-edging or a strictly edge moving? One allows a hole, the other not. In addition to this problem, we have the infinity of the plane, so strictly speaking, keeping the edge in itself, is impossible. Is the dominance of $C_n \circ P^n$ enough to replace the idea of the edge moving in itself? Clearly not. It only corresponds to the edge ending up on the edge! We can easily find infinite transformation that keeps the far away points far away and still will leave a hole in the values.

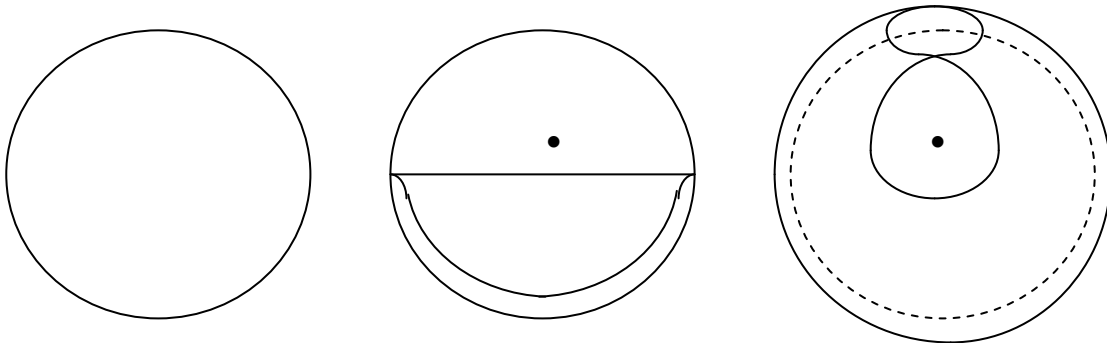
Easiest to achieve this, is by creating a hole in a unit circle, as on a table, and then blow out to infinity, the whole transformation.

So why are the algebraic $\mathcal{A}(P)$ transformations strictly inside edge motions and never folded ones, allowing holes. This is the real question, not why do we have roots.

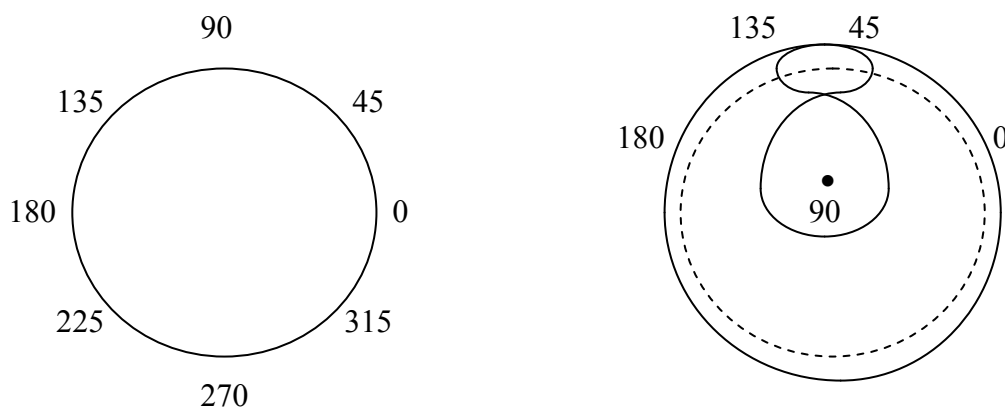
If $\mathcal{A}(P)$ picks up all points, it picks up the origin O too, so we have root obviously.

The explanations are already hidden in the finite table situation!

It is not true that the edge folded motions end up the same as the turnings with the edge in itself. Clearly, when we fold the tablecloth, some points suddenly are outside the edge. Then we stretch the half edge to cover the whole circle, leaving a hole.

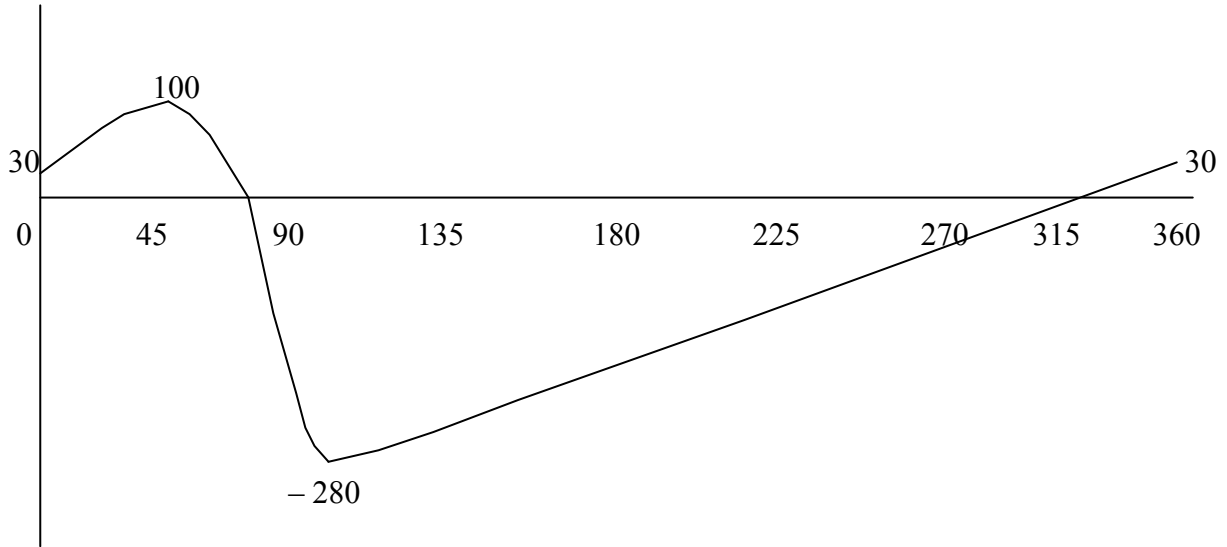


Looking from these left out points in the hole, our effort to cover the edge looks quite different if we now introduce a new time aspect, namely the “looking around”. In fact, this can be even better seen if the original sheet had angle numbers around it.

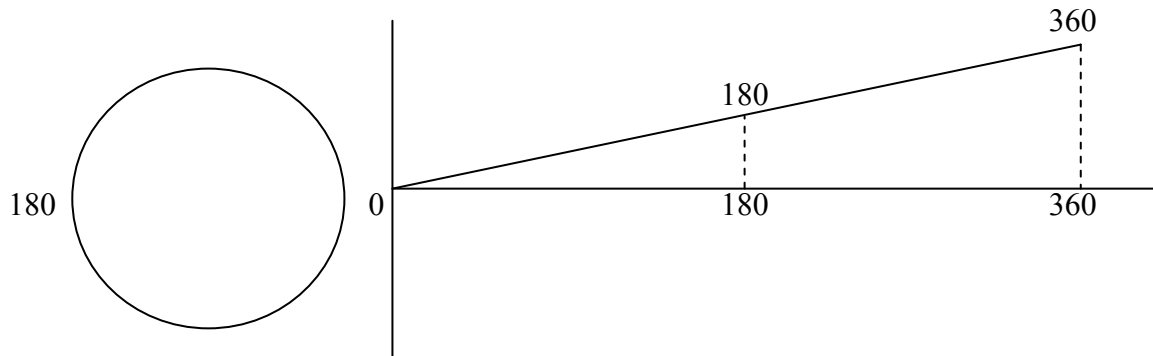


If the left out observer follows the numbers in this “cover up”, then a strange double circling is happening in his actual angles with two reversals in the direction.

This can be seen even better by graphing the angles of view as a function of the original angles.



It doesn't look that strange, but lets compare it with a real full circle without any distortion:

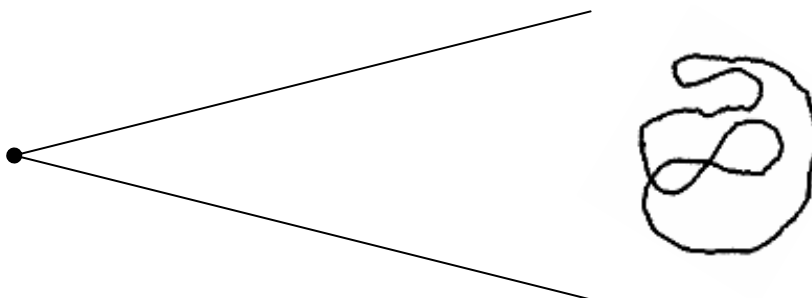


It's not the proportions that are important, in fact, the view angle was in much smaller scale than the original angles. The vital fact is that here, the view started from 0 but ended on 360° . These two angles are the same in the plane as measurements, but not as the track of the journey. Above, the view angle actually returned to the 30° . So in fact, we didn't do a circle at all!

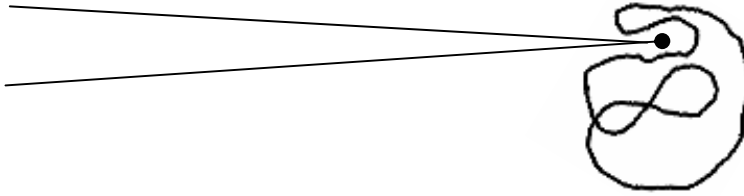
Most amazingly, actual circling of a point can be even more complicated than these two.

We can go around more than 360° and still return to accomplish nothing! On the other hand, we can also circle round and round with more than one full 360° above or under the initial angle. In short, the view angle can end up to be $\pm m 360$ to the initial.

If we are far away from a loop then we can be sure that we are outside of it, because it fits into our view with a much less than 360° angle. So it doesn't surround us, even subjectively.



A less trivial case is when the “escape” angle is small, but still guarantees outsideness.



Unreal surrounding by the angles is much harder to distinguish from real:

From the five points in this picture:

One is obviously outside.

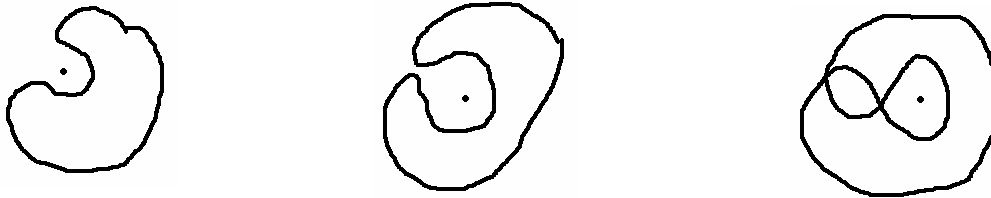
One is “outside” by the angles.

And three are inside because they have surrounding loop with 360° total.



The crucial fact is the following:

If an observer is fix, but a loop is continually moving in the plane, then it can go from obvious outsider to phony surround by the angles. But it can never change into a surround by the angles, unless it touches the observer point.



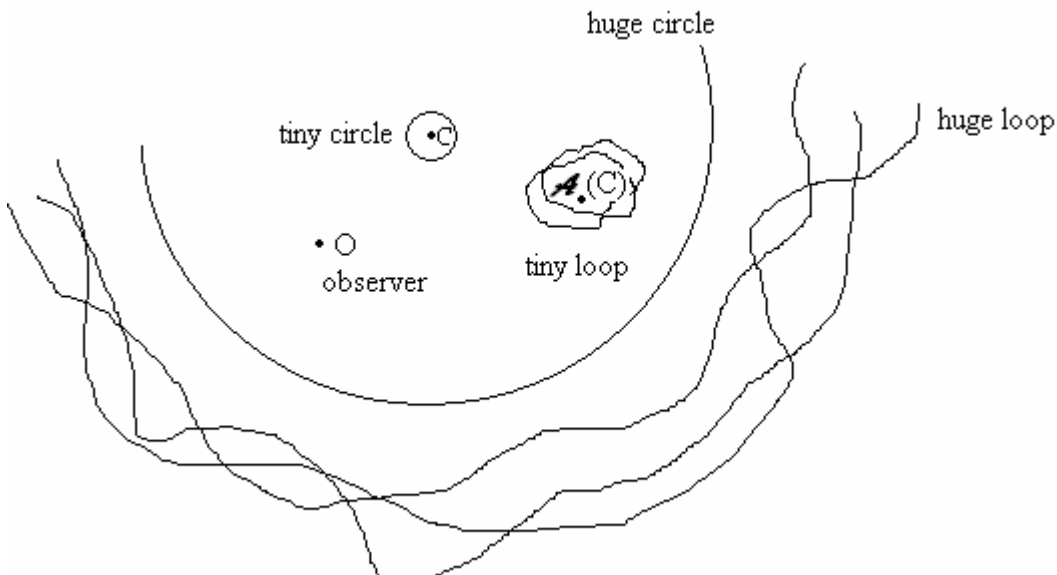
Now the domination of the $C_n \circ P^n$ member can show its additional feature!

It is not only becoming further than any fix observer if P increases, but as P goes around once, $C_n \circ P^n$ will go around n times. In short, as P increases in circles, actual surrounding will happen. That's why no hole can appear. The precise argument goes as follows:

Suppose O were an outside point, that is not taken up as $O = \mathcal{A}(P)$.

Then lets choose any C point and a small enough circle around it, so that its image by \mathcal{A} is such a tiny loop, that O is outside of it obviously with an escape angle.

Now increase the tiny circle continually! Its image eventually increases too. In fact, after a while, the circle is big enough that its loop image by \mathcal{A} , will definitely surround O , namely with $n \cdot 360^\circ$ total angle:



But a continually changing, non surrounding loop cannot become a surrounder, unless it passes through the observation point.

Here, this could not happen, because O was assumed to be not a value.

History will never reveal how much of all this was clear to Gauss because he decided not to reveal it. He did return to newer and newer proofs for this so called Fundamental Theorem of Algebra. But he was chasing more an illusion of exactness, than a bigger picture for others. When he turned his attention from the plane as new continuous numbers, to the grids as new naturals, there was another chance to avoid the formalism of $\sqrt{-1}$.

The Vultures of Formalism

Gauss, not stripping away the $\sqrt{-1}$ form and reveal the geometrical meaning behind what he called the $x + y i$ complex numbers as merely (x, y) points for the Fundamental Theorem of Algebra, and grids for the complex integers, was a didactical laziness. He simply didn't care about over-analyzing the vision. He was onto something much bigger. He was chasing a new truth. He was talking about the new numbers as a new layer in the world of shadows. He was physical. His idol, Newton, was something very special in human history. This is beyond being one of the greatest minds and yet one of the most disturbed geniuses at the same time. The scientific uniqueness of Newton is that he was the first and last mathematician physicist. Conventional science historians would again try and disprove me by bringing up formal counter examples about other scientists crossing over. But the simple truth is that being a mathematician is a category and being a physicist is another. Understanding of course, is open to both fields, in fact for everybody. But the truth is even more shocking! Before Newton, there were no physicists. Researching reality, going from alchemy to chemistry, measuring water and air pressure are all very nice things, but have nothing to do with the mind frame that a Maxwell, Einstein, Heisenberg possessed. The crucial shift of course is most detectable looking right before Newton. Two minds were at the verge of crossing over. Galileo, this wandering, sweet and stubborn rebel, and Kepler, a lost mathematician. If they had paid more attention to each other, they could have crossed over. Usually, when Newton's revolution is explained, we go into how forces and gravitation especially combine the earthly and heavenly worlds of Galileo and Kepler, blah blah blah. But all these slide over something much deeper. A hidden new role of mathematics. Kepler was the first to inject math into nature. Before him, the math of physics was too simple to notice it. He, using Tycho Brahe's data collection could only hit upon the idea of ellipses, because he was a closet mathematician. Then claiming that God put the planets on ellipses, was where he missed to be a physicist. Newton's deriving of the Kepler laws from the laws of forces, was merely the first grand offering on a new altar. Newton re-defined God. After Newton, anybody believing in white thrones or lotus flowers as the actual realities of the other side, is bound to be an ignorant and false witness. Math does not have to be injected into nature! Math is behind reality or math is the deeper reality! To play with this deeper reality is a freeing of one's spirit. And indeed, many became mathematicians throughout the ages. But to become a physicist, is a whole new "game". It is much younger than becoming a mathematician, and so "we" don't have a correct picture of it yet. By "we", here I meant philosophers, the only ones who care to see the biggest picture. So all these complicated thoughts come back to the explanation, why Gauss didn't care about explaining and clarifying the visions. He was a secret, closet physicist. Instead of feeling the immediate solution to his pain, he gazed into the future. Where complex numbers are used in Quantum Mechanics. He saw something far away. That's the usual cause of not seeing what's right in front of our eyes. We'll come back to this crucial blind spot in Gauss' life, but now I want to talk about the vultures. This is the real reason why Gauss' and all other geniuses fault in not caring enough to explain can have devastating effect on history. The army of epigones who teach the teachers not to teach, create a cesspool of rotten food for the minds. Common sense of course prevails as the rejection of the meaningless curriculums. Kids hate math and potential geniuses rot in jails. The silver platter concepts slowly do penetrate the system of lies, but school should be an education of these silver platter concepts. Using them and attacking them. None of these are happening in schools today.

Since mathematics is playing in the garden of truth, it is easy to stray in it. To require math as a separate subject is actually the trick of the devil, to lock people out from its paradise. Physics, which is a much newer mystery, can not be abused as much as math. So it is simply locked out from the education itself. In physics, the lies can not be piled up with the same formal derivations as requisites. So the fact that people don't possess the minimal vision of physical reality is left unquestioned. It just melts into the technological second nature that we live with but live without in spirit. Very few mathematicians come to this strange, final analysis. In fact, I don't know anyone besides me, who is a champion of New Math and at the same time, knows that math shouldn't be a subject on its own. Only integrated science is correct didactically.

The morsels of common sense must be guided to abstractions. Abstractions cannot be taught on their own. To teach all kinds of complicated math to kids, who can't even visualize the atmospheric pressure, is an obscenity. But most sadly it is an intentional conspiracy of the devil. So we arrived from Gauss not caring, to a world misleading its children. Once somebody loses sight of the big picture, then the veil of spell will lead him to his particular truth, which can still point to the right directions. This book is no different! It approaches something very simple, in a very round about way. Falling and climbing again. Up until now, it seemed like a math book. We'll see! Accepting math, even as a subject is a lie, as I just said. I already mentioned about the changing methods in the teaching of second order equations, and how ridiculous is to solve it by guessing roots from the coefficients. Other whole fields, like calculus is going through cyclic re-evaluation by the narrow minded vultures. Calculus in, calculus out. Every ten years, they re-play this game. As slowly the old dogmas of trigonometry fade, and new "hip" fields like probability, are introduced with the same dogmatic approach, the teachers are trained to obey, so the students have no chance in hell. The stupidity of math on its own, is almost at the point of revealing itself. Forty years ago, it was argued whether a common high school curriculum is meaningful at all. They said, "Lets do some more real education towards jobs, not just for potential tertiary levels.". Against this I could say that schools should be windows to the universe, not doors to the world. But that's beside the point now, when 90% of kids who learn math are doing it just to get into medical school and never to use it anyway. The rottenness of education is beyond repair, nobody believes that school can make you better. It is the test to be bad enough for the world. To do didactically correct lines has a new meaning which I am approaching right now. There was a time when it already become clear why education is a lie.

Timothy Leary said, "Turn on, tune in, drop out.". Without going into the depth and clarity of this, just observe that "drop out" was the end of it. No didactical analysis was required to see that. But then, for a short time, it seemed that instead of the "knowledge is power" the "seeing is freedom" can be true. The fact that this polarizing of the attitudes could die off, and become an "era", a potential "been there, done that", already shows that the trouble is much deeper than anybody can guess. The taste of the word "establishment" still remains, so hundreds of years from now, Timothy Leary will receive his place in the philosophical hierarchy, but the truth of the sixties is blocked by the beast it tried to unlock. No knockouts in this fight, not even scores. The mediocre nimrods who saw nothing more than unwashed rebels didn't win. The Nixons, the Reagens, the Bushes, the conservative stupidity of "honest day's work", cannot even put a finger on the pulse. Their own doesn't exist. Zombies or skeletons, as Ginsberg called them. Their focusing is self induced narrowing of the iris! Anything against the establishment painted as dangers against freedom. The puppets of establishments are hypnotized, but there is no conspiracy, there are no puppeteers. The Beast is bigger than Man and it doesn't need men to enforce its will. The truth of the lies must come out in even more tangible ways, undeniable, irrefutable. Common sense nonsense is the goal. When all details of the emperor's clothes are adored, can it only become transparent. So it's all good. The ex-hippies are surfing the net, and Wikipedia can replace schools. The stupidity is the same and the illusion of freedom is stronger. Before, the libraries were the overwhelming collections of useless details. Now they are the sane refuge from the bewildering emptiness of the internet.

Where will this book end up? In libraries or on the net? I don't know it yet.

Right now I have to climb again to look behind the simple proof we just created by the plane curves. So we are to unveil the "delusions of density" and "confusions of continuity":

32. Density at points

An S set of points is dense at P , or S is approaching P , or P is a limit point of S , if: In every surrounding of P , there are infinite many S elements.

Such P can be two kinds, element of S or not.

Both can have a special case when a surrounding is fully in S :

P is inner point if all points of a surrounding around it are S elements.

P is a hole if all points of a surrounding around it are S elements except P itself.

The density points that are not inner or hole can be called mixing points, because every surrounding of these contain infinite many S elements and infinite many non S elements.

These mixing points can be still elements or not.

The density points or limit points of S are abbreviated as $\lim S$.

The S elements of these, that is, $S \cap \lim S$ is abbreviated as $S' =$ dense part of S .

S is dense in itself, or in short, dense, if all S elements are density points, that is approached, that is limits, that is $S \subseteq \lim S$, that is $S' = S$.

S is closed if $\lim S \subseteq S$, that is there are no approached non elements.

The inner points of S is (S) . S is open if all elements are inner, that is, $S = (S)$.

This seems to be a stupid name, but the justification for it is that if \bar{S} denotes the complements of S , containing the non elements of S , then:

$$S \text{ is closed} \leftrightarrow \bar{S} \text{ is open}$$

The only open and closed set is the full space, and formally, the empty set \emptyset .

Indeed, being closed means containing all density points, but if there are no points, there are no densities. Also, S is open if all points are inner, but if there are no points, then they are "all" inner. The rule itself is quite obvious, because an S is closed, if and only if all non elements are not approached by elements. That is, all non elements are inner in \bar{S} .

33. Density in an interval

Intervals are of special importance. In space they mean cubes. A surrounding is the same on the line, while in space it is a ball. But cube or ball, could be both used instead of each other.

Even the word, "where", will refer to not points, rather to intervals. The "somewhere" means in some interval, "nowhere" in no interval, and "everywhere" is in every interval.

Of course, these are only meaningful if we first define something for intervals by its points. Simplest is "fullness". The I interval is fully in S , or I is full in S , or S is full in I , if $I \subseteq S$.

So nowhere full then means that S has no full I in it. The next most important is density:

S is dense in I if S is dense at all points of I . So S approaches all points of I or $I \subseteq \lim S$.

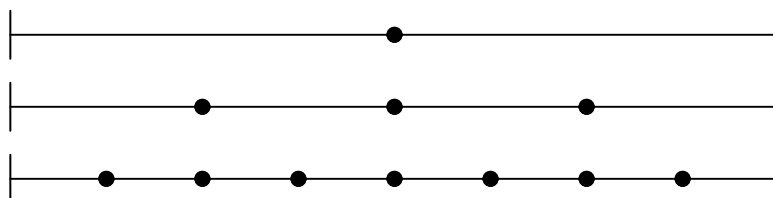
Again, a nowhere dense S is not dense in any interval. This of course implies that S is nowhere full too. The nowhere dense is also called by me as "rare".

The visual image of a rare or nowhere dense S should be the following:

Every I must have an I_0 sub-interval, where S has no element at all, so I_0 is empty of S .

Indeed, obviously if such I_0 window exists, then S can't be dense in I , because for example, the center of I_0 is not approached. But in reverse too, if every I_0 sub-interval of an I would contain S element, then every $P \in I$ can be approached by such sub-intervals, containing points of S and thus, S would approach P too.

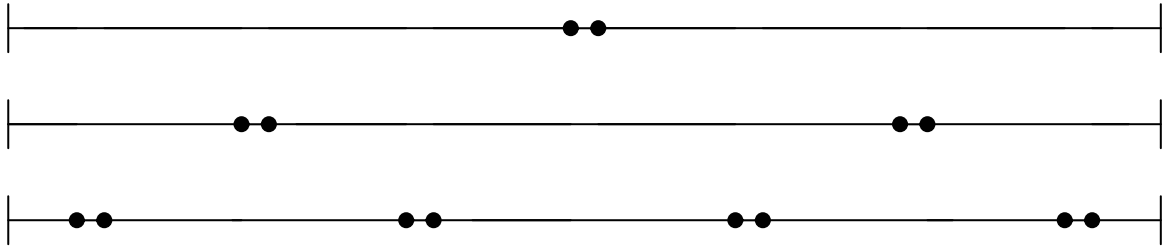
The simplest set dense in I is the repeated halving points of I :



They obviously will approach every point of I . This includes themselves.

So, the halfers are dense in itself, or in short, dense. This might give the impression that being dense in itself implies being dense in I . The counter example is simple:

Instead of single halfers, lets use double points or twins:



Of course, eventually, we have to use closer and closer twins, but we never put new points between the twins. Thus, they indeed remain twins, that is they are the next points to each other on their left or right. From the other side, of course, other twins will approach them.

So at once, it's clear that all elements of S are approached. It is dense in itself. In short, dense. Yet, this set is not dense in any I , including the full in which we created it.

Indeed, the twins enter every I , no matter how small one we choose. The twins will go there, and even inside, thus taking their little window with them. So indeed, these windows are I_0 sub-intervals without S points.

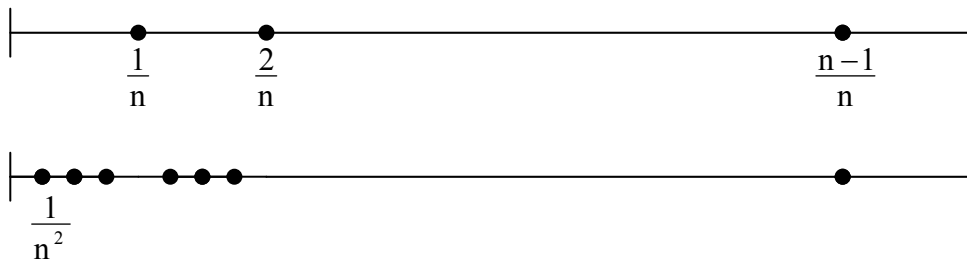
Thus, S is a dense yet rare set.

34. Decimals and n -dividers

By showing the amazing twins, we slid over a hidden problem of the simple halfers.

They are clearly dense in I , so dense in itself too. But are they the same density points? In short, how do we know if there are points of I that are not halfers at all? Well, from simple fractions we know that $\frac{1}{3}$ can not be expressed as halves or quarters or eights and so on.

The generalization of the halfers are the n -dividers.



This is exactly the decimal system for $n = 10$.

Even the 10-dividers, that is the finite decimals can't give all fractions. In fact, the mentioned $\frac{1}{3}$ is only an infinite approached point: $.333333333 \dots$

The division algorithm tells us that all fractions are cyclic or periodic. So this gives the incredible knowledge that even all dividers can't create all the points of an I .

Namely, the non periodic infinite decimals are definitely not dividers.

The Greeks didn't have this silver platter and only realized by geometric constructions, that some distances can't be dividers, that is fractions. The silver platters of the infinite decimals and the division process of course, still don't show us where these non dividers are.

To see them, we have to tell more about "how they are".

35. Sequencing the dividers, closed non dividers

The fix n-dividers are very easy to list by going to deeper and deeper divisions. For example, the 10-dividers or finite decimals are:

$.1, .2, \dots, .9, .01, .02, \dots, .99, .001, .002, \dots, .999, \dots$

first division
second division
third division

The obvious, but strange fact is that all these up to a point are repeated among the finer divisions. For example, $.3 = .30 = .300 \dots$. This gives a good idea:

Why don't we just list the first divisions, but for all n-dividers, by increasing n:

$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \dots$

Indeed, the finer divisions of 5 for example, will come anyway at $n = 25 = 5^2$.

So we just listed all dividers or fractions of 1.

This could be called the third silver platter for points.

It provides us with an even more amazing version of the twins.

We know that among the dividers are the non dividers, but now we'll pick twins from these in an amazing way with the help of the sequenced dividers:

First, pick two non dividers, so that $\frac{1}{2}$ is between them. Clearly, a lot of other dividers will

fall between them too, but, leaving those out from the sequence, it will still contain infinite many. Now lets find the first in the remaining sequence that is left and the first that is right, from our twins. The first in the remaining sequence of course itself must be one of these.

Now for these two left and right dividers, we can again create two non divider twins, containing the dividers between them. There will be other dividers of course in between, which can be left out from the sequence. Now we can choose four new first dividers to be left and right from the previous two. And most importantly, the first in the sequence must be one of these. And so on.

The sequence of dividers must empty out, because we always used the first, after leaving out some. The resulting sequence of twins is a set S made of non dividers and is a dense rare set as before. But now it contains all dividers between the twins! So now, it's even more puzzling where the twins are. But if the S twins are puzzling, then $\lim S$ is even more so.

Clearly, $S \subseteq \lim S$, because all twins are approached by twins themselves, but are there $\lim S$ points outside S ? One thing is sure, those limits can't be between the twins, and so, can't be dividers. So, the $\lim S$ set is closed and is outside the dividers or fractions or rationals.

It is a closed subset of the irrationals!

36. Isolation

Lets return to point sets in space, in general. The opposite of an approached or density point is a P that has a surrounding with only finite many S elements inside. Then, there must be a closest to P and so, taking a small enough surrounding, we have only two possibilities:

If $P \notin S$, then there are no S elements inside the surrounding at all, so P is an outer point.

If $P \in S$, then it is the only S element inside, so P is an isolated element of S .

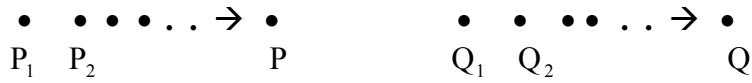
The isolated points of S is the S_0 isolated part of S .

$S = S_0 \cup S'$ is the splitting of S into its isolated and dense parts.

S is isolated if $S = S_0$ that is, $S' = \emptyset$.

37. Densing the density

Leaving out the S_0 isolated points from an S , that is keeping S' , we'll only have density points of S , but they may not all be density of S' itself. In short, new isolated points of S' can appear. For example,



$$S = \{P_1, P_2, \dots, P, Q_1, Q_2, \dots, Q\}$$

$$S_0 = \{P_1, P_2, \dots, Q_1, Q_2, \dots\}$$

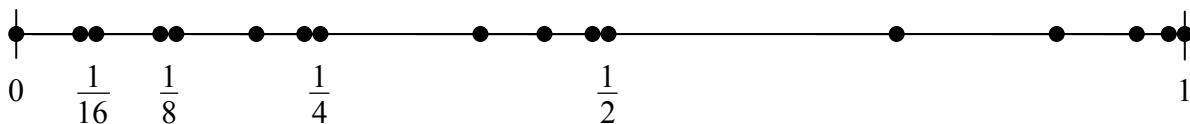
$$S' = \{P, Q\}$$

Both P, Q elements of S' are isolated in S' . The more complicated the S set is, the further we can go in leaving out the isolated points from S, S', S'', S''', \dots and thus, form the next ones. But even an infinite sequence may not finish the job.

Luckily, there is a way to obtain the “strong density” points of S directly too. The clue is this: If S is dense, then all S points are strong dense too, because there are no isolated points to leave out at all. Now, if a T subset of S is dense, then also this T can only contain strong density points of S too. Indeed, it can't contain isolated P point of S , because it were isolated in T too. Similarly, the higher isolated ones were also included in T . What's more, if we have more such T dense subsets, then their total is again dense. Indeed, if $T_1 \cup T_2 \cup \dots$ had an isolated P , then it were already isolated in the T_i that contained P . So the total of all dense subsets of S is the widest dense subset, containing exactly the strong dense points of S . This total is S^* , the strong dense part of S . Remember that S^* is actually $S'''''' \dots$ in a precise way. The $S - S^*$ is the thin or scattered part of S . This split of $S = (S - S^*) \cup S^*$ is more important than the earlier $S = S_0 \cup S'$. For dense S sets, that is when $S_0 = \emptyset$ of course, $S = S' = S^*$.

38. Zero sets

S^* is a smart but abstract way to capture $S'''''' \dots$. There is a very visual consequence when $S^* = S'''''' \dots = \emptyset$. We showed a picture with two $P_1, P_2, \dots \rightarrow P$ and $Q_1, Q_2, \dots \rightarrow Q$ limits in S . Obviously, there can be infinite many. In fact, we can put them in an I interval, say $[0, 1]$ easily, if the limits are $1, \frac{1}{2}, \frac{1}{4}, \dots$



So the limits themselves have a limit, namely 0 . Now, $S' = \{1, \frac{1}{2}, \frac{1}{4}, \dots, 0\}$ and so $S'' = \{0\}$. The amazing is this: If we cover 0 with a small I_0 interval, then only finite many limits are uncovered. Then, covering these with I_1, I_2, \dots, I_n finite many points are left out at all. They can be covered with I_{n+1}, \dots, I_N . So, all together, arbitrary small and finite many $I_0, I_1, I_2, \dots, I_N$ covered the whole S set.

Having higher $S'''' \dots$ would go the same way. If they die out, that is $S^* = \emptyset$, then finite many arbitrary small intervals can cover the whole S set.

Such finitely and arbitrary small coverable sets are called zero sets.

Though our ideas that lead to the zero sets were good, we made two errors.

Firstly, the limits don't have to be in S . Luckily, $\lim S$ is closed. Indeed every limit of limits is already limit of S . Thus, $(\lim S)^* = \emptyset$ implies that S is zero. This a stronger condition than $S^* = \emptyset$ because $S^* \subseteq (\lim S)^*$. Indeed, any T dense subset of S is also of $\lim S$.

Namely: T is dense subset of $S \rightarrow T$ is subset of $S' \rightarrow T$ is subset of $\lim S$.

But even $(\lim S)^* = \emptyset$ is not enough because our second mistake was that we only looked at sets in an I . Some sets can not be contained in an interval, that is they are not bounded.

Simplest is the naturals as an S set of points:

$\bullet \quad \bullet \quad \bullet \quad \bullet \quad \dots$
 1 2 3

$\lim S = \emptyset$, so $(\lim S)^* = \emptyset$ too, but S is not zero. We can't cover it with finite many pieces. Indeed, the fundamental truth is:

Zero sets are bounded and rare.

The boundedness is trivial, because any I_1, I_2, \dots, I_m is bounded.

For the rareness, lets assume that S is coverable by finite many arbitrary small intervals and I is any interval. Then S is also coverable by $\frac{I}{2}$ totalling I_1, I_2, \dots, I_m closed intervals.

Finite length additions cannot increase, so obviously I_1, I_2, \dots, I_m cannot cover I , that is some $P \in I$ is not covered. But, $I_1 \cup I_2 \cup \dots \cup I_m$ is a closed set too, so all non elements of it are outer. Thus, P is outer point of $I_1 \cup I_2 \cup \dots \cup I_m$ and so also outer of S , and so indeed, P is not approached by S . Thus, S is not dense in I .

Now that we are so precise, we should also re-examine the original idea. $(\lim S)'''' \dots$ dying out, leads to the finite many points and thus, S being zero. The crucial element is that:

If S is bounded and $\lim S = \emptyset$ then S is finite.

In negative form, Bolzano-Weierstrass theorem:

If S is bounded and infinite, then S is dense at some point.

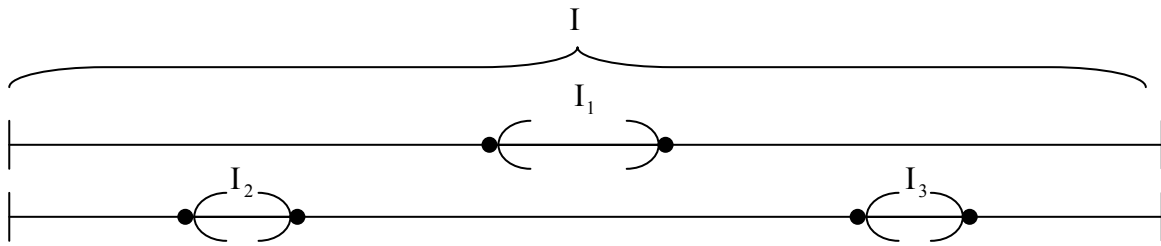
Indeed, let S be bounded by I . Since S has infinite many points in I , at least one half of I must have infinite many points too. We can halve this again, and so on, we find narrowing halving intervals, that all have infinite many S elements inside. The limit or common point of these intervals has to be a limit of S too. This hidden assumption that the narrowing intervals themselves have a common point will be examined soon in more detail.

39. Twin width, twin splitters

Since all zero sets are bounded and rare, a good question is if there are bounded and rare non zero sets. Also, since all bounded $(\lim S)^* = \emptyset$ sets are zero, another good question is whether there are bounded $(\lim S)^* \neq \emptyset$ zero sets. This of course boils down to finding zero set that $\lim S$ is dense or already S is dense. So, since the twins are bounded, dense and rare, they are definitely examples for one of these questions. If they can be both non zero and zero then they are examples for both.

We already used the crucial fact of the zero cover, that it can be chosen closed and so keeps the limit. Thus, if some twins are zero set, then their mysterious $\lim S$ can be included, that is

$S \cup \lim S = \lim S$ is zero too. Even though we don't see "where" $\lim S$ goes out of S , a big advantage of accepting the full $\lim S$ is that then we can simply omit the obviously non limit points, namely the windows between the twins.



In fact, we can calculate the remaining $\lim S$ set's size as:

$$I - I_1 - I_2 - I_3 - \dots = I - (I_1 + I_2 + I_3 + \dots)$$

Since we made the windows, we can make them from any I_0 sub-interval of I .

The distribution is easy by using smaller and smaller halves, that is $I_0 = \frac{I_0}{2} \cup \frac{I_0}{4} \cup \dots$

But, we also have to cut these for the increasing doubling pieces. So:

$$I_1 = \frac{I_0}{2}$$

$$I_2 = I_3 = \frac{1}{2} \cdot \frac{I_0}{4} = \frac{I_0}{8}$$

$$I_4 = I_5 = I_6 = I_7 = \frac{1}{4} \cdot \frac{I_0}{8} = \frac{I_0}{32}$$

$$I_8 = \dots = I_{15} = \frac{1}{8} \cdot \frac{I_0}{16} = \frac{I_0}{128} \quad \text{and so on.}$$

Now if $I_0 < I$ then $I - I_0 > 0$, so since the windows are I_0 in total, the $\lim S$ is $I - I_0 > 0$, so can't be a zero set. Thus, these twins can't be zero either.

This interval size argument of course, relied on the assumption that I_0 can't cover I .

We'll come back to this too. Another interesting question is whether $I_0 = I$ is possible.

That is, using the full I for windows, can we even obtain a leftover $\lim S$ set? Perfectly!

In fact, then the zeroness of S is directly seeable. As more windows are cut out of I , the leftover becomes arbitrary small and finite many pieces can cover it. Such $I_0 = I$ windowed

twins could be called "wide" twins. But we don't have to start widely with $I_1 = \frac{I}{2}$, it's only the tendency that counts.

Now lets place one-one point between all twins anywhere in the windows!

These "twin splitters" S are obviously an isolated set, $S_0 = S$. Thus $S' = \emptyset$, so $S^* = \emptyset$ too.

But they approach all the twins (and their limits too). So $(\lim S)^* \neq \emptyset$.

So, the twin splitters are examples why we had to refine our arguments from S to $\lim S$.

Most importantly, the twin splitters are always non zero!

Indeed, first of all if the twins are not wide, then we can widen them keeping the splitters. Then, the twins are a zero set and so the left or right half of them is zero too. Using the splitters as alternative twins for a kept half of the original twins, we gat narrow twins. These are not zero!

But a total of two zero sets is again zero! So if the splitters were zero then the narrow twins had to be too.

40. The boundary problem

Usually I don't give a damn about accepted namings if they are illogical. In fact, for a while, I went overboard and created too complicated names because I wanted them to reflect their meanings. But even though I don't accept that "a name is just a name", I had to accept that too long or strange names are no practical. So I compromised. Density itself was something that I wanted to banish, due to its delusional feature. But then, I embraced it even more and I think it's a very faithful name, if we know that it means similar yet different things, like dense in itself or dense in a full interval. In fact, I went even further and accepted rare for the nowhere dense. Another case was the closed-open sets. The closed name is obviously alright because it reflects that the limits belong, so are not leading out, so indeed are "closed". But having only inner points, could also be regarded as a certain "closedness" of the points. Of course, the mentioned complementarity duality makes closed-open perfect. Strangely, this seemingly compromising acceptance of open sets as open made a new meaning too to me by the "boundary problem". I didn't introduce the name boundary, because it didn't serve any purpose yet. It is a continuation of the closed-open complementarity naming.

Boundary points are the mixing, isolated, and hole points. So in negative, all points, except the inner or outer. First it seems stupid to include isolated or hole in a "boundary", but observe that the boundary of an S is exactly the same as the boundary of \bar{S} . Indeed:

An isolated of S is hole of \bar{S} .

A hole of S is isolated of \bar{S} .

A mixing of S is mixing of \bar{S} too.

So the boundary is indeed the boundary between S and \bar{S} , belongs to both.

This boundary is thus also, a "skin", but with the weird inclusion of isolated and holes.

The open sets are without their boundary or skin, the mixing points. So indeed, they are "open" to the world, like "infections".

In fact, this whole line of closed-open and boundary naming has a very deep layer that in fact relates to where we are going, namely continuity.

Are physical objects point sets?

A physicist would protest. After all, the atoms are jumps in matter and inside them, everything seems different. But, is our everyday object vision usable for point sets?

Unfortunately, not without problems.

The basic dilemma is whether an object should be closed or open.

Is there a last surface belonging to the object?

If so, then putting a book on top of another, couldn't be visualized.

Since there are no points next to each other in space, thus, the top of the lower book and the bottom layer of the top book, couldn't be next to each other.

We could say that a thin layer of open air is between them.

A much better idea is to regard objects as open sets without the surface. Then this could be explained as the force field of the outer most atoms. So then the books on each other, are pushing into each other's force field, due to the force of gravitation.

The first atomists were of course the Greeks and they were pondering about points, space and time a lot too. I already mentioned many times the silver platter expression, referring to a kind of easy solution for hard problems. Of course, most of these were not easy to find at all. The division of numbers in a base system, the infinite decimals, the coordinate system, sequencing. They all had to wait thousands of years. The "easy" rather refers to the understanding, namely that these are so natural to use. But I also mentioned that this "easy" is again just a trap. They leave undigested thoughts, they don't really fix up all the loose ends. All in all, the silver platters are useful and good. They melt into a bigger set of evolution of archetypes, the same way as the language of cinema changes our imagination for the better or worse.

41. Continuity

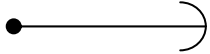
Now this is a problem so deep and hard to even raise that it is the mother of all others.

How are the points located in space, how is time flowing as moments of “nows”?

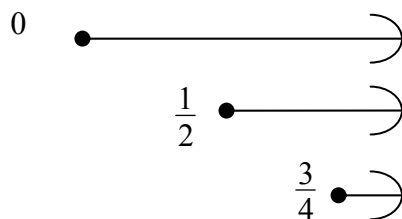
I already mentioned the Achilles paradox and how it is today trivialized by our decimal system.

But the Greeks went deeper, they realized that without strange competitions, just a flying arrow in itself, is hard to swallow. Going too deep, we might miss the crucial layers that are closer to the surface. Nobody knows this better than the mathematicians, some of whom were openly anti philosophical. But this was merely because they never grasped real philosophy.

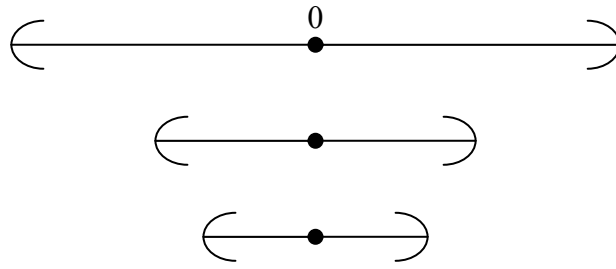
With continuity, the silver platter or “near surface” solution, was formalized by Dedekind.

He was one of the few people who at once realized Cantor’s genius and full heartedly followed him. The Dedekind axiom of the line is so simple, so embarrassingly obvious, that we can’t believe it was under the surface for so long. It says that if we cut a line into two, the left and right side, then exactly one of them must have an “end”, a “closure” of the “cutting wound”, while the other remains open, that is without end point. So it is a more specific version of what I already mentioned above about points not being next to each other. This not next to each other feature was known before Dedekind and Cantor. Already the Greeks realized that the points must be dense on the line, so any two must have other points in between. Now a Greek philosopher might say to Dedekind: “How do you cut a line?”. And especially my visual portrayal of “wound” would be unacceptable. Especially, because we can’t remove the two halves of a cut line. The real meaning of Dedekind’s cut is merely dividing the points to the left and right side. Now this can’t be argued against! The total of the points of the line is a set. These elements all have the left, right relations to each other. So its merely by logic that such distribution or separation or “cutting” must exist. And then Dedekind tells the simple law, how the two halves must be. As they say, the test of the pudding is the eating, so the test of Dedekind’s ingenious axiom is using it. To derive other things from it. The best test is to derive the much more complicated principle of Cantor, the master, who was thus perfectionized by his follower. Cantor said, that if we zoom in or narrow down with intervals, then they don’t just disappear, rather will contain a common point. Now if we visualize a point, and then the intervals that contain this, in other words we go backwards, then this common pointness is so trivial that we even doubt if it can say anything. But lets visualize only what it says! Intervals inside each other! So they can dance back and forth, the second can be almost at the right end of the first, then the third at the left end of the second and so on. We might even think that such dancing loses hold of the final point. No way. That’s exactly the “point”. There has to be a common point. In fact, if we don’t require that the intervals become arbitrary small, then we can have more common points. And this is the main issue: Having common point. The single common point is merely a consequence of the size shrinking, and has nothing to do with them being in each other. And yet we run into problems here. First of all, we can present a counter example: Lets use intervals that have left end but not right end, like $[0, 1)$: 

Now narrow down with: $[\frac{1}{2}, 1)$, $[\frac{3}{4}, 1)$, . . .



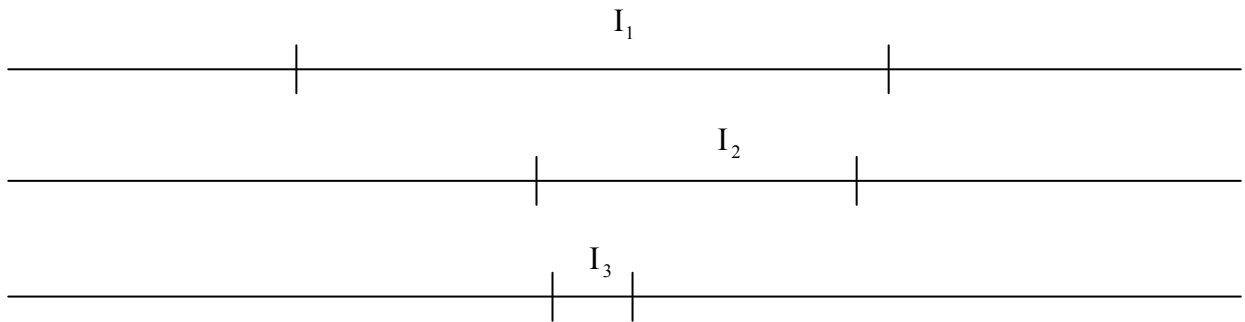
As we see, we narrow down to the point 1 , but it won’t be in. Okay, so we have to use close intervals! But that’s not right either, because even both end open intervals can zoom into a common point, for example 0 by: $(-1, 1)$, $(-\frac{1}{2}, \frac{1}{2})$, $(-\frac{1}{3}, \frac{1}{3})$, . . .



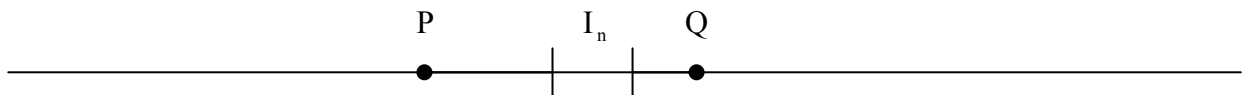
In fact, we could zoom to any $(-a, b)$ interval instead of 0 and thus, contain all in between points. So it seems as a messy affair. But only seemingly! The main fact is that closed intervals nested must have at least one common point. The rest is just particulars.

So can we derive this from Dedekind?

Let the nested intervals be $I_1 \supseteq I_2 \supseteq I_3 \supseteq \dots$



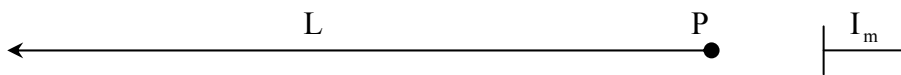
Now let L be those points of the line that are left from any I_n and R be those points that are right of any I_n . We might think that these are the left and right cuts of Dedekind, but we are wrong. It would be too easy. Indeed, L is left from R , that is much obvious, because if $P \in L$ then P was left from an I_m , so it is also left from all I_{m+1}, I_{m+2}, \dots while if a $Q \in R$, then it was right from $I_n, I_{n+1}, I_{n+2}, \dots$. Now which ever is bigger m or n , that will be in both of these intervals, so if it was for example I_n , then:



So P is left from Q too.

But these L, R don't have to be the full line, in fact, they never are.

Another thing they can't be is the halves that have an end point. Indeed, if for example, L had a right end point P then P would be the right most point and still left from an I_m :

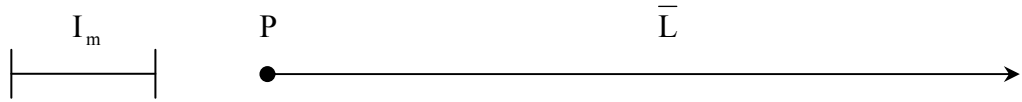


Which of course is impossible, because between P and the left end of I_m , we have points also left from I_m . So L has no end. There is no last point that is left from some I_m .

Now lets regard the real right side of L , that is the \bar{L} complement from the full line.

This \bar{L} will then have a left end point P by Dedekind's axiom.

But what points are in \bar{L} ? Well, all those that don't belong to L , that is points that are not left from any I_m . We claim that P is in every I_m ! Clearly, P is not left from any I_m , so it could only be either in I_m or right from it. But it can't be right from it, because then it couldn't be the left end of these.



Indeed, between the right end of I_m and P , the points are also not left to any other I_n either, no matter whether I_m is in I_n or I_n is in I_m .

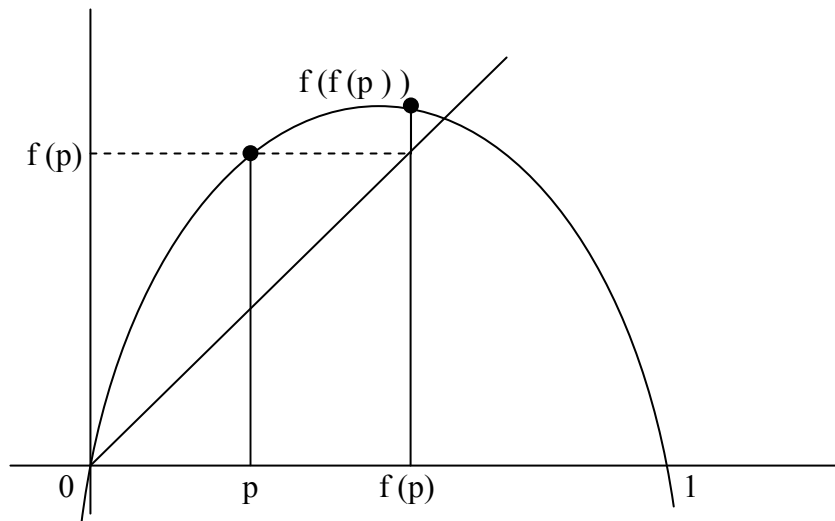
So indeed, the left end of L is a common point. Of course, similarly, we could have used the right end of \bar{R} . Quite an elaborate logic! Which explains why Dedekind's axiom is so simple and Cantor's so complicated.

42. Iteration, an embarrassing ignorance

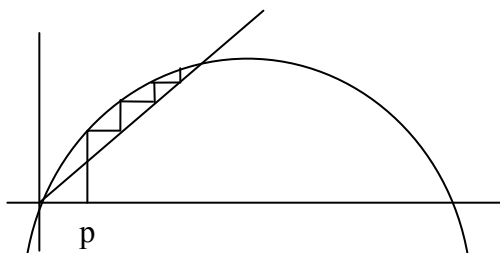
First there was Calculus, discovered by Newton and Leibniz. It slowly became Analysis, with higher precision, namely not relying on subjective continuity assumptions. Finally, Cantor and Dedekind turned all that into Set Theory and Topology. The question remained whether the continuity axiom tells everything about the subjective continuity. It seems that it tells everything that is inside continuity, but it definitely didn't tell everything that we relate to continuity. Namely, an arrogant assumption was that if a value changes continually, then everything that we calculate from that value, will also change continually. For simple calculations, this is so. But if the p value is used as a starting value for an f function that we repeatedly apply infinitely, that is regard $f(p)$, $f(f(p))$, $f(f(f(p)))$, \dots then even for simple f functions, the full infinite sequence can be very non continuously depending on p .

The simplest case is $f(p) = rp(1-p) = r(p-p^2)$. This even has a biological meaning!

The p value is the population density changing from 0 to 1. The r constant is the reproduction and food shortage combined factor. And f is the yearly or periodic population change. Indeed, big p is advantageous because more animals breed more. But too many will have less food and opportunity to mate. So, $p(1-p)$ is a proper dependence. The f function is really simple! An upside down parabola over $[0, 1]$. The iteration even has a beautiful geometrical picture. Indeed, to use $f(p)$ as the new p value, we simply have to draw a horizontal line to the 45° line and look for the new parabola value above or under this:



Earlier, people overlooked this whole affair, because most of the time, it is indeed boringly obvious. For example, if the parabola is flatter, f merely climbs up to its newer and newer values:



But if the bump is there before the 45° line crosses the parabola, then weird “back and forth cycles” and even stranger things can and will happen. The most important though is that if the r constant is right for these, then the p starting value is crucial too. In fact, changing p continually will only cause continuous segments, but not be continuous in the full $[0, 1]$.

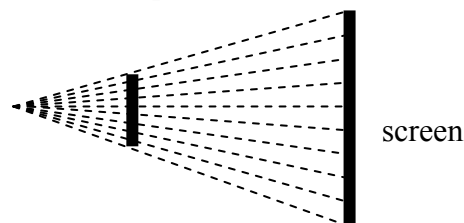
The biological example is not ideal for this discontinuity, after all a population is only a discreet value in reality anyway. But, the geometrical meaning shows that other mirrorings, reflections, behave the same way. So for example, a magic lamp would change non continuously in the dark as a bright and dark light show. A less important side feature in this example is certain repeataces in the continuous segments. The two dimensional fractal versions of these iterations emphasize this not exactly self repeating feature as the emergence of infinitely deep patterns.

The mathematics of these are not very deep, but the fact that they remained unrecognized till the 20th century is very surprising. The rise of computers of course, helped a lot in recognizing the patterns, but it doesn't explain how the fundamental potential existence was missed earlier.

43. Continuum

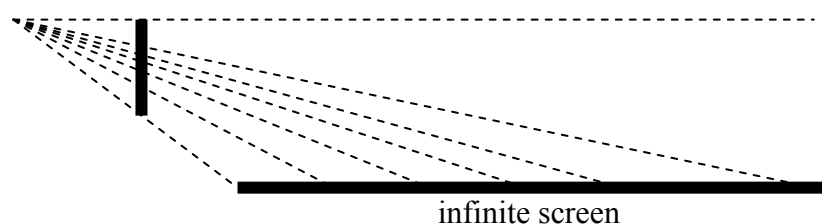
As opposed to these trendy fields like Chaos Theory, Bifurcation, Fractals, the other word relating to continuity, the “continuum”, remained a mere curiosity. Here of course, we have a very misleading name, which I nevertheless keep, because its falsity is reflective on a bigger mystery. Linguistically, there is nothing wrong. While continuity as noun, comes from the adjective continuous and means its essence, continuum is a noun too, but refers to the “bearer”, that is the continuous sets. So intervals, line, square, cube, circle, ball, triangles, . . . are all continuums. But it's not their continuity that is important in giving such common name for them. To cut right into the problem, lets present a paradox. In fact, the biggest paradox to me is why this wasn't raised by the Greeks. Especially because it relates to the Euclideaness of space. As I explained, it means that proportionally enlarging pictures, the angles don't change, we don't have distortion. A sweet little kitten, becomes a huge “little” kitten, and not a mutant freak.

In real life projection, the small film square or rectangle contains all the dots that are projected to the screen. If we assume infinity of points, then nothing really changes:



But this means, that the small film has same many points as the screen. In one sense, it's even logical, because it would just mean that infinity is absolute. But the size of the screen being big, is still a fact. Actually, putting a film onto the screen, the whole thing becomes much more puzzling. Indeed, we then project a tiny bit of the film onto itself. In real life, the same can happen, but then the quality drops. Here, a perfect point by point blowing into itself happens.

At any rate, this projection shows that the continuums or at least some of them, have same many points. Amazingly, even an infinite line can be projected too:



Not only the Euclideaness was immaterial in these projections, but even the Dedekind or Cantor continuity was not explicitly used. The fact that size doesn't matter, still leaves the question of whether dimension does? That is, a square or cube would be the same many points as a line.

44. The whole-part paradox of infinities

The projection paradox is not unique. In fact, all infinities have subsets that are exactly the same as the whole. The simplest example is the shifting of the naturals:

1	2	3	4	5	6	7	8	
				→	1	2	3	4

Starting from 5 we have the same sequence, or to put it another way, we can delete the beginning and still have the same left.

When Galileo formulated his law of consecutive falling distances as the odd multiples, he observed that:

1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8

So even half of an infinite is the same.

At projections of course, the paradox was much stronger, but now it seems that we get sub-identities for any sets. We can push the strangeness a bit even further. If the half is the same, then the half of the half is too and then again, so it seems we can find even infinite many sub-sequences in a sequence. Unfortunately, we made a mistake. As we use half of the halves repeatedly, they empty out, so we can only do the trick finite many times. But let's not give up.

A much simpler trick works:

1	2	4	7	11
	3	5	8	12
		6	9	13
			10	14
				15

Each of these sequences have different numbers in them, that is, they are disjoint.

Their total is the set of all naturals. And they are all the same infinity.

The big dilemma is this. Does this feature of the infinities having so many same subsets make their comparison meaningless? Both the projection and the sequencing or lining up, has the fundamental idea of ordering to each other elements like labels or names or twins. But is this something tangible or an illusion?

45. Comparison

The simplest argument for believing in this twin ordering or so called equivalence, can strangely be obtained by forgetting about infinities. Suppose we have a huge ballroom with boys and girls. We want to know which is more. We could count them both, but it takes a long time. Instead, all we have to do is tell them to form pairs. Indeed, if they do that, then whichever is left without pairs, is the more. Strangely, the reason this works is exactly the opposite feature of finites to infinities. They can't be equivalent to their subset. In fact, if another set is equivalent to a subset of one, then this must be more. So sadly, this argument made us even more doubtful, if infinities can be compared by equivalence. Indeed, they can be equivalent to their own subset, so another infinite being so, doesn't mean squat either, yes? Wrong!

Let's take a closer look at the finite case. Being paired to a subset of the girls means that the boys can't be more. They were physically paired to a subset of the girls! The verdict that therefore, the girls are more, was a second phase, a logical conclusion, namely that the reverse, that is pairing all the girls to boys is impossible. We merely derived this conclusion, because we knew that we only had finite many of them in the ballroom. So the solution is simple.

With infinities, we need again two phases. The first is to order every element of S to some elements of T . This is the physical part and it guarantees at once that T is at least as much as

S , or that S can't be more than T . In short, $S \leq T$.

The second phase must again be a logical conclusion from some feature of S and T .

Namely, that perfect ordering of the S elements to all T is impossible, that is, $S \not\sim T$.

What if such perfect $S \sim T$ is impossible, but the T can also be ordered to an S subset, so $T \leq S$ would be? Luckily, this is impossible. It can be fairly easily shown that:

If $S \leq T$ and $T \leq S$ then there is perfect pairing too, that is $S \sim T$.

But there is a more fundamental problem. What if the easy part, that is $S \leq T$ is established and we simply can't find $S \sim T$ because it doesn't exist, but we can't prove this, because the features of S, T don't imply it. Well, then tough luck. Then we are in a limbo.

The $S < T$ claim that S is smaller than T , means that $S \leq T$, so the physical part must be true, and we claim that $S \sim T$ is impossible. So then this impossibility is meant by logical means. This is especially convenient, because our logic is indirect. So all proofs boil down to getting contradictions from false assumptions. In short, to prove $S < T$, we have to show a contradiction in $S \sim T$. Since we aim for a contradiction of $S \sim T$, we might as well assume $S \sim T$ and get a contradiction using this too. So in short, the prototype of a proof is:

$$S \sim T \rightarrow S \not\sim T$$

In a more concrete sense, it means that we assume $S \sim T$ and then we still find that some S elements were after all left unordered to T elements. This left out unordered part of S might even just be a single element which seems negligible in view of infinities. But we have to see that it was a merely a logical act, showing the $S \sim T$ impossibility.

The above mentioned problem of $S \sim T$ not following from S, T features, but neither leading to contradiction, is a real issue and the root of some deep mysteries later.

46. The three levels of point infinities

The smallest subjective infinity of course is a P_1, P_2, \dots sequence of points.

The first grand result is:

The continuum is not sequencable.

So let I be an interval and P_1, P_2, \dots a sequence of points in I .

The physical part is thus done:

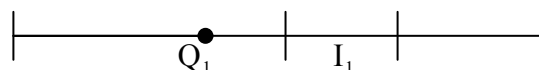
$$\{P_1, P_2, \dots\} \leq I$$

Now we have to show that we can't have equivalence, that is:

$$\{P_1, P_2, \dots\} \sim \{Q_1, Q_2, \dots\} = I$$

Here, the Q_1 is the pair ordered to P_1 then Q_2 to P_2 , and so on.

But, Q_1, Q_2, \dots is again a sequence, so it's enough to show that a sequence can't be the total I . Choose any I_1 closed interval in I , that doesn't contain Q_1 .



Now, Q_2 may be in I_1 or outside. Either way, we can find again an I_2 in I_1 that doesn't contain Q_2 . Then in I_2 an I_3 that doesn't contain Q_3 , and so on.

I_1, I_2, \dots must have at least one common point, say P . This can't be any of the Q_1, Q_2, \dots because they are each outside of I_1, I_2, \dots

As I said, the chased contradiction seems very narrow! A single measly missing P . But this is misleading. There have to be more missing points than the double, triple, even infinite many repeats of Q_1, Q_2, \dots .

The real question is whether we can give better pictures where these more missing points are. Yes, we'll do that later.

Now we'll go above even the continuum. We find sets that are bigger than all the points of an I :

The point sets of I are more than just the points.

The easy part is again trivial. Every point is itself a point set by regarding it as a singlet set. So all we have to show is that:

$$I \neq \{S; S \subseteq I\}$$

A false copy of our previous proof would be the following:

Here again we have the trivial subset as the smaller set, so again we only have to show that there is a missing subset. This is trivial too, because not all subsets are singlets.

This was too easy (and stupid)! Above the reason for the missing point search, was not the subsetness, rather that the equivalence of a sequence is again a sequence.

The naturals are used hidden as subscripts!

Here I is the total interval. Ordering all of its points to some subsets, these are not a collection of singlet subsets anymore! So now, we have to assume that any $P \in I$ has an S_p point set ordered to it, and then show that some S has to be missing from the S_p -s.

For a P , the chosen S_p either does contain P itself or not. Lets collect all those P , where S_p doesn't contain P and that will be the missing S . That is, $S = \{P; P \notin S_p\}$.

Now why is this missing from the S_p -s? Well suppose it were an S_Q !

Then again, either $Q \in S_Q$ or $Q \notin S_Q$. Contradictingly, both would imply the other:

$Q \in S_Q$ means $Q \in \{P; P \notin S_p\}$, so $Q \notin S_Q$.

$Q \notin S_Q$ means $Q \notin \{P; P \notin S_p\}$, so $Q \in S_Q$.

Observe that $S = \{P; P \in S_p\}$ wouldn't lead to contradiction. Indeed, if this S is S_Q , then:

$Q \in S_Q$ means $Q \in \{P; P \in S_p\}$, so $Q \in S_Q$ which was the assumption anyway

$Q \notin S_Q$ means $Q \notin \{P; P \in S_p\}$, so $Q \notin S_Q$ again.

This whole argument for the subsets of I didn't use anything about points.

So it is a universal fact that there are more subsets of any set than elements.

We'll go toward this universal purely set feature later, but now we are interested in points.

So lets return to the continuum being more than a sequence.

First of all, this gives a new and purest proof for the fact that the dividers, that is the fractions, are not all the points. As I mentioned earlier, we have the division process and the infinite decimals as silver platters that convincingly show how the non fractions or non rationals, that is, irrationals have to be much more than the irrationals. Indeed, there have to be more totally random infinite decimals, than periodic ones. But I also said, that these silver platters still don't show how all that can be on the line. How it is embedded into continuity.

Well, the contradictory equivalence showing that the points of I as a set is more than a sequence, is even less helpful as the decimals. It is in fact, the purest, least detailed argument.

But we can use such pure arguments to achieve even more surprises about continuity.

47. Sequencing all fractions

We already sequenced the dividers or simple fractions as increasing first level dividers:

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \dots$$

The set of all fractions is an obvious sequence of sequences:

$$\begin{array}{cccccc} \frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{4}{1} & \frac{5}{1} & \dots \\ \frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \dots \\ \frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \frac{5}{3} & \dots \\ \frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \frac{5}{4} & \dots \\ \cdot & & & & & \cdot \end{array}$$

We need a reverse of what we already showed, how to split a single sequence into infinite many:

$$\begin{array}{cccccc} 1 & 2 & 4 & 7 & 10 & \dots \\ & 3 & 5 & 8 & 11 & \dots \\ & & 6 & 9 & 12 & \dots \\ & & & 10 & 13 & \dots \end{array}$$

And indeed, a sequence of sequences can be pushed ahead the same way, line by line:

$$\begin{array}{cccccc} \frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \frac{4}{1} & \frac{5}{1} & \dots \\ & \frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \frac{4}{2} & \frac{5}{2} & \dots \\ & & \frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \frac{4}{3} & \frac{5}{3} & \dots \\ & & & \frac{1}{4} & \frac{2}{4} & \frac{3}{4} & \frac{4}{4} & \frac{5}{4} & \dots \end{array}$$

And then, we can sequence it from left to right, column by column:

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{3}{1}, \frac{2}{2}, \frac{1}{3}, \frac{4}{1}, \dots$$

Thus, the set of all fractions can be sequenced. They are a dense set in the whole positive half line, and yet, they must be smaller than any continuum, thus even the tiniest full interval.

48. The greatest concept of Mathematics

I said that the non sequencability of the continuum is the purest way to prove that there are points that are not dividers. Our proof for this unsequencability used Cantor's common point principle, so relied on the continuity of the interval. So the name, continuum, made perfect sense. Now we'll change all that. We'll show that continuity is not the real reason that the interval is not sequencable. To be more precise, continuity is merely a background condition.

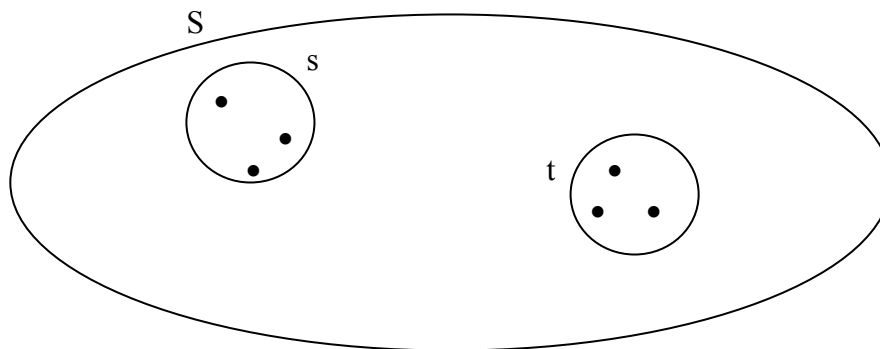
To be even more precise, a set doesn't have to be continuous, that is doesn't even have to contain any full interval to be non sequencable. So, with our earlier "some", "no", "every" agreements, a nowhere full set can be non sequencable too. In fact, we'll show that even a nowhere dense set can be non sequencable. This might suggest that we simply obtain a new non sequencability. Sets that are more than a sequence, but less than the full interval. We wish that would be the case, but it isn't. It will be evident that this new property that guarantees the non sequencability, stands for the full I too. In fact, this new property is so pure and independent from points and intervals, that it will even encompass the seemingly most abstract proof we used for the next infinities, the point sets or subsets of I .

So what is this magical new concept? Well, it is Choice!

To show how heuristic is this concept, let's do the just mentioned transformation, that is going from subsets to choices. The solution is very simple. If every element of an S set has a choice of two words, "yes" or "no", and we pick one of these double choices for every element, then we get exactly the subsets. Indeed, all we have to do is regard all those elements in a total choice set, that had one of the possible choices, for example the yes. If every choice set was formed, then every subset comes about. This is even more obvious if instead of yes/no, we use in/out. Then, the "ins" are indeed, the elements that are in an S_0 subset. The "outs" will of course define the $S - S_0$ complementers of S_0 .

This opens up a whole new door. Why should we choose from only two choices? We can even use different many choices for the different elements. Two of course is the minimum. One choice is not a choice. And that's the whole point. If we do have at least two choices for every element of an S set, then the possible choices always must be more than the elements, that is S . Here of course, possible choices, meant simultaneous choices for all elements, that is a choice set. Or rather a choice function. Indeed, an f choice function gives for every $s \in S$ the $f(s)$ choice for s . Of course, function is itself, just a set, namely, the set of all pairs $(s, f(s))$.

The real importance is that s must be included in the simultaneousness, in order to differentiate. If for example, f is a yes/no or in/out dual choice, and we wouldn't attach s to its choice, then a choice set would melt or collapse into the yes, no duo. But in theory, it could be that all the s elements have their own choices, different from what other t elements can have:



Here we drew s, t as themselves being sets and their elements are the possible choices.

Then we don't have to attach s to its choices and t to its. So a choice could be merely a sample from the S set, that is picking one element from each of its elements.

These alternative reformulations are still the same. The essence is the choice!

The choice as the magical wand of Set Theory, became infamous, not for this heuristic feature that it creates a bigger set, rather quite oppositely as a method to create an equal sized set in another one. That is, to compare any two sets at least to the first physical phase. Subjectively it means, that we pick one-one element from both sets S and T , say s and t . We order these to each other as a beginning of equivalence. Then from $S - s$ and $T - t$ we pick new elements that we order to each other, leave them out and pick new elements and so on. Whichever of the S, T sets disappears or empties out first, will be thus equivalent to a subset of the other.

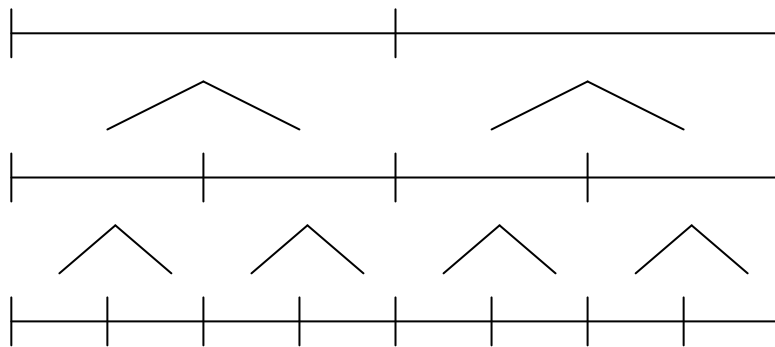
This whole thing sounds a bit loose, relying too much on time as the next picks.

The grand story of early Set Theory was that time could be eliminated completely. We can make pre-determined choices for all possible continuations. Then the existence of a total choice set is a provable theorem. So the comparability of sets doesn't have to be an axiom, it is a theorem. But the grand proof of this theorem uses choice sets that are not like the fix yes/no. They are not from a fix pre-determined set, rather from the possible leftovers. It is still totally physical, but as it turned out, Logic, on its own cannot claim that at least one full choice set exists. So its strange, but true:

If a set has elements, then Logic can derive that we can pick an element, but to do so simultaneously for a set of sets, that is to obtain a sample, or picking function has to be stated as a new axiom. This recognition was the deepest moment of mathematics yet. And the resulting Axiom of Choice is the deepest axiom of mathematics! A link between Logic and Set Theory.

What followed became the most stupid misrepresentation of truth. All problems that surfaced in Set Theory were falsely reduced to this purest and most reliable axiom. We can understand that among all this circus, this other importance of the choices as size increasers, wasn't even mentioned. Lets see how this choice picture works for the full I first.

The halving sequence now is viewed as not of points rather as left or right half intervals.



Now every "end" point of the interval is a point not only determined by the narrowing halving sequence of intervals, but also by the sequence of left/right choices.

The dividers themselves are only approached by sequences that become all left or all right from a point. All other, that is alternating sequences will determine some mysterious, not visible points, depending on every single choice we make in the lefts and rights.

For dividers, like in base 10 decimals, everything is the same and now we see a bit more of what's behind the silver platter. Every infinite decimal number is a choice sequence with the ten digits as possible choices. But it doesn't make any difference how many choices we have, if we have at least two. Why? Well lets see for the decimal choices.

Suppose all decimals, that is choice sequences, could be sequenced as:

1	. 3 7 0 4 9 2 4 9 4 1 5 . . .
2	. 4 0 5 6 1 2 4 9 3 2 1 . . .
3	. 2 3 7 9 4 6 1 8 7 6 4 . . .
.	
.	

Now lets regard the D diagonal decimal that is made from the first digit of the first, second of the second and so on:

$D = . 3 0 7 . . .$ This could be anywhere in our list. But lets change every digit of D , so make an "anti diagonal" \bar{D} , for example by adding 1 to the digits, meaning 0 from 9.

$\bar{D} = . 4 1 8 . . .$ This \bar{D} can't be in the list! Indeed, it's not the first, because the first digit is definitely not correct. It's not the second, because the second is definitely not correct there. And so on. It can't be any of them. So \bar{D} is missing from the list.

49. The continuum loses its meaning. Dense limit

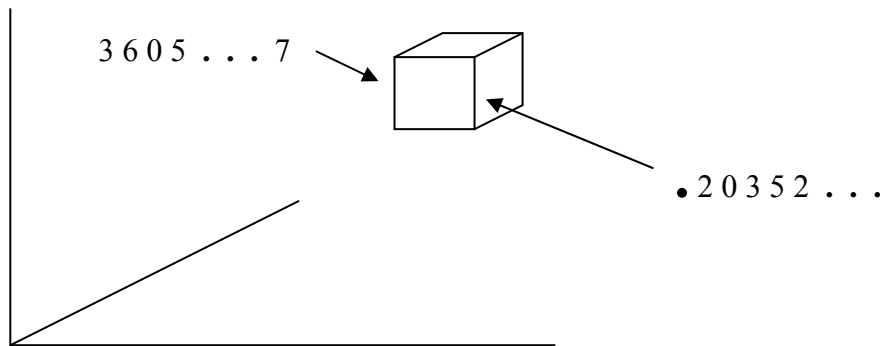
Now we are ready to show why the continuity is not the essence of the continuum. We still keep the word continuum for this size as a reminder of where it came from. Its proper name should be choice sequence size.

T Dense set's limit is continuum, that is, $S \subseteq \lim S \rightarrow \lim S \sim \text{continuum}$

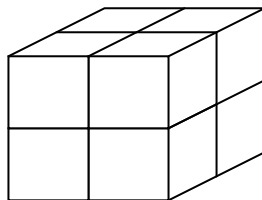
P I want to give the proof in three dimension, where we don't rely on the left and right ordering of the line. So first, we need to show that the full space is a continuum.

The projection of an interval to the half line shows that the infinite decimals with a whole part are also continuum. So it's enough to show that all points of space can be also marked by such

3 6 0 5 . . . 7 • 2 0 3 5 2 . . . infinite decimals. We need two tricks. The whole part will tell in what unit cube the point is inside, then the decimal part will tell the location in the unit cube:



The unit cubes of the full space can be easily sequenced by starting from the origin and using bigger and bigger unit layers.



This layer 1 has 8 cubes. Layer 2, which is putting cubes around layer 1, will have $4^3 - 8 = 56$ unit cubes. Then layer 3 will have $6^3 - 4^3$, and so on. In the layers, we can list the cubes in any order, but going through the increasing layers, we sequence all unit cubes.

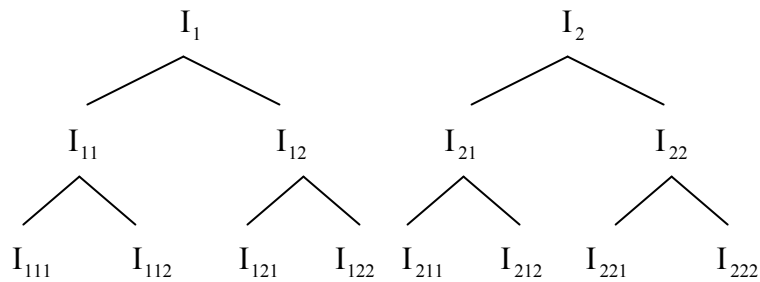
Now comes the second part. Using the decimals of the x, y, z coordinates in the unit cube, we can comb them together into a single decimal, telling the x, y, z point. For example:

(.2 5 9 6 0 3 . . . , .0 1 2 5 3 9 7 . . . , .6 1 7 9 1 3 . . .) is ordered to
 .2 0 6 5 1 1 9 2 7 6 3 9 0 9 1 3 7 3 . . .

Now that we established that the whole space is a continuum, it's enough to show that dense set's limit must contain a subset that is already continuum too. Indeed, then the total limit set can't be neither bigger, nor smaller either.

We'll create a sequence of pairs of cubes nested in each other, giving a sequence of choices for the actual nestings and each leading to a point of $\lim S$. The first I_1, I_2 cubes can be arbitrary but disjoint and each containing S points inside too, not only on their surface.

Then, in both of them, must be I_{11} , I_{12} and I_{21} , I_{22} respectively, again disjoint and each containing S points inside:



Since all intervals contain S points, thus any P determined by a narrowing choice sequence will be approached by S . The density of S guarantees that there are always infinite many points inside the intervals, because an inside point can only be approached inside.

For the half dividers, the theorem simply gives what we already proved, that I is continuum.

But for the S twins, we get that the limit points, $\lim S$ is a continuum too.

In short, we obtained a rare continuum.

50. The real continuum problem

Above in 40, where I introduced the new concept of choice, and I mentioned that we might think it leads to a new non sequencability, I said, “We wish that would be the case.”.

That was a reference to this real continuum problem.

When Cantor realized that not just the full interval, but rare sets can be continuum, he thought that the reason for this is that simply there is no room for other infinity. There is no infinity between the sequence and the continuum. But as time passed, and nobody could prove this claim, the focus shifted. People were chasing weird point sets that would be in between.

It was similar to the road to non Euclidean geometry. That started also as a nagging question about the parallelity axiom and then the chasing of a contradiction lead to a new possibility.

But here, quite oppositely, no new world appeared!

Cantor’s original belief became known as the Continuum Hypothesis. The generalization of it is that for any S set too, the infinity of choices from S is the next infinity after S ’s own.

This also couldn’t be proved or disproved with even one counter example. So is there a point even in looking at points? Maybe it’s all just some Set Theoretical conspiracy. That’s why I think it is appropriate to keep the word continuum instead of the “sequence choice”. Because, even if it turns out that the point sets are an illusion behind this puzzle, then even more the continuity is a puzzle, how it can appear so real. So the name continuum will remember this.

In fact, Paul Cohen proved that from the axioms of Set Theory, these Continuum Hypotheses are not decidable. So we can neither prove nor disprove that between S and its choices, there are infinities. For the general case, we might respond cynically about Set Theory as a whole.

But point sets on a single interval cannot be thrown out as over complicated fictions.

Thus, we continue on with our investigations of the continuity and the continuum.

51. A “closed case”

Lets recall how the trivial $S = S_0 \cup S'$ splitting of any set into its isolated and dense points was improved into an $S = (S - S^*) \cup S^*$ split with S^* being the strongly dense part. Intuitively, this S^* is $S'' \dots$, that is repeated densing, that is repeatedly omitting the isolated points. But, S^* can be directly defined too as the widest dense subset of S . So it is dense. This is the big improvement from S' , which is not necessarily dense, it can contain isolated points of its own. So now that we showed that a dense limit is continuum, the denseness of S^* becomes important. Of course, only the limit is continuum, that is $\lim S^*$, not S^* itself. So we can’t jump to the conclusion that S is continuum too, since it contains S^* .

And indeed, such conclusion would mean that all sets are continuum, which is absurd. But there was a formal error in this conclusion too, namely that S^* can be empty, so even $\lim S^*$ is not continuum. The $S^* = \emptyset$ sets are those that contain no dense subset at all, that is all subsets contain isolated point. If $\lim S \subseteq S$, that is S contains all of its limit points, that is S is closed, then $\lim S^* \subseteq S$ too, and thus indeed, S must be continuum, if S^* exists at all.

By the way, if S is closed, then a dense subset of S doesn't have to be closed too, but the widest S^* has to. Indeed, if a P limit point of S^* were outside S^* , then adding P to S^* , we would obtain a wider dense subset. So for closed S , $\lim S^* \subseteq S^*$.

This way the continuumness of S is even clearer:

- 1.) Closed dense set is continuum, since the limit is inside.
- 2.) Any S set that has closed dense subset is also continuum.
- 3.) Any closed S that has dense subset, has closed dense subset too, namely the widest dense subset S^* is such. And so, S must be continuum.

Two questions remain:

- 1.) If $S^* = \emptyset$, how big is S ?
- 2.) If $S^* \neq \emptyset$, how big is the scattered part $S - S^*$?

The answer is the same, sequencable.

1.) means that a closed S is either continuum (if $S^* \neq \emptyset$) or sequencable (if $S^* = \emptyset$). So among closed sets, the continuum hypothesis stands.

2.) is merely an interesting detail, that closed continuums have their split also by size too.

Luckily, we don't have to prove 1.) and 2.) separately, because the common feature that guarantees the sequencability is that S or $S - S^*$ doesn't contain dense subset.

This means that every subset has isolated point!

T
P

If every subset of S has isolated point, then S is sequencable.

We can regard again the isolated points of S , that is S_0 . Leave them out, that is take $S' = S - S_0$. Then leave out the isolated ones again, and so on. We obtain the full S as:

$$S = S_0 \cup \underbrace{(S - S_0)_0}_{S'_0} \cup \underbrace{(S - S_0 - S'_0)_0}_{S''_0} \cup \dots$$

$$\underbrace{\hspace{15em}}_{S'}$$

In spite of this seemingly complicated form, the important fact is that the first isolation, is the main one. S_0 is "more" than S' ! Namely, all S' points are approached by S_0 ones.

Indeed, if a $P \in S'$ were not approached by S_0 , then in a surrounding of P , there were only S' elements. By our assumption, that all subsets have isolated points, the S' elements in this bubble should have one P_0 too. But then, this P_0 were not only isolated in the bubble, that is in S' , but also in the whole S and thus, P_0 were in S_0 , which is outside the bubble.

This fact that all S' points are approached by S_0 points, means only subjectively, that S_0 is more than S' . Indeed, if we would want to actually, pair a distinctive S_0 point to each approached S' point, then we were in trouble! It's true that there are infinite many approaching S_0 points to choose from, but different S' points are approached by sequences that share elements. We feel that from the infinity of choices, we can pick distinctive ones too, but this has to be demonstrated. The trick is to aim higher.

We're going to order to every $P \in S'$ a whole sequence of S_0 points, distinctly!

This will be an $F(P)$ function. Picking one from $F(P)$ as $f(P)$ of course then gives an equivalence from S' to a subset of S_0 . The real trick is that this abstract F, f will be built by G, g through the isolated parts of S' .

Starting with S'_0 giving G is easy. Indeed, each point in S'_0 can have different bubbles and so disjoint approaching sequences from S_0 . Then pick one from each and this is g .

The S'' points are approached by S'_0 points and so obviously by G points too, but also by g points alone. So we can extend G to S''_0 again by the bubbles, from g . Picking again one-one element, we extend g too. The g values of course will be a mere subset of the previous ones. Continuing the same way, in the end, $G - g$ will be our perfect F .

So indeed, S_0 is "much more" than S' . All S' points can be labelled by unique infinite many S_0 points. In the cold world of equivalences of course, it merely means that S' can't be more than S_0 . More usefully, if we can show that S_0 is sequencable, then $S_0 \cup S' = S$ is too.

The sequencing of S_0 is again a bit surprising. Clearly, every point in S_0 can have the already mentioned disjoint bubbles around them. How many bubbles can be placed in the whole space? They are like actual balls, some big exercise ones, some little ping pong balls, and some can become arbitrary small dust balls. Filling them tightly in space, it's quite a lot, so the sequencing is not evident at all. Every little box in space can have infinite many in it.

We already saw that density in the infinite is misleading because the fractions of the whole half line is merely a sequence. The rational points of the plane then are merely a sequence of sequence too, because for every rational x we have infinite many rational y . Then we can go to space, by regarding a new sequence of the sequences in planes, by fixing z for each rational value. But we can directly give a sequencing of all rational triplets, by the increasing totals of the coordinate numerators and denominators:

The smallest value is 3 for the origin as $(\frac{0}{1}, \frac{0}{1}, \frac{0}{1})$. The next value is 4, which can be:

$(\frac{1}{1}, \frac{0}{1}, \frac{0}{1})$ or $(\frac{0}{1}, \frac{1}{1}, \frac{0}{1})$ or $(\frac{0}{1}, \frac{0}{1}, \frac{1}{1})$ or new variants of the origin as:

$(\frac{0}{2}, \frac{0}{1}, \frac{0}{1})$ or $(\frac{0}{1}, \frac{0}{2}, \frac{0}{1})$ or $(\frac{0}{1}, \frac{0}{1}, \frac{0}{2})$, so six triplets all together.

The next 5 value allows much more combinations: $(\frac{1}{1}, \frac{1}{1}, \frac{0}{1})$, still finite many.

And so on, each total allows finite many triplets, so we can sequence all triplets as we go.

Now that all fractional points are sequenced in the positive quadrant of space, it's obvious that they can be in the whole space too, by going through the eight quadrants for each triplet.

Finally, the bubbles must all contain rational points, so picking one for each, we can sequence the bubbles and thus, the points of P_0 too.

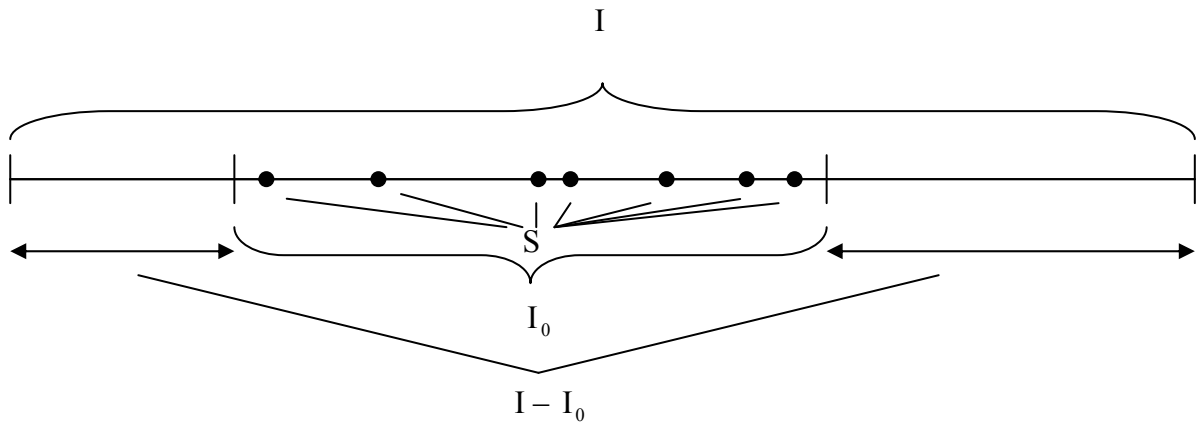
52. Coverings

The closed sets being either continuum or sequence was a little too complicated to follow.

The complementing open sets on the contrary are trivially all continuum, because they contain full surroundings. This strange complementing simplicity versus complicatedness, is telling something. Containing a full surrounding or interval at once implies continuumness, but the whole point of the choice picture was that it showed continuum dense limits, which don't have to contain full interval. In fact, they can be rare.

There is a much simpler direct argument to achieve continuum, using complementarity.

For example, if an S set in I can be covered by using only the I_0 part of I , then this not only means that S is smaller than I_0 , but that $I - S$ is bigger than $I - I_0$:



As we see, $I - I_0$ comes together from two pieces, but it is obviously the same as $I - I_0$ were, by cutting off I_0 from I .

So then, we might as well cut I_0 into pieces, that is use it more economically, to cover S .

Still, it has to be true that if I_0 can cover S then $I - S$ is at least as big as $I - I_0$.

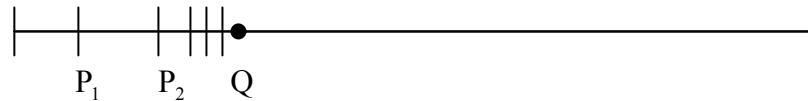
So then, even though $I - S$ doesn't have to contain a full interval, it is bigger than any smaller than $I - I_0$ interval, and so, $I - S$ should be continuum.

Thus, we obtained a new method of creating continuums.

Of course, there are a lot of loose ends here:

How do we cut the I_0 into pieces?

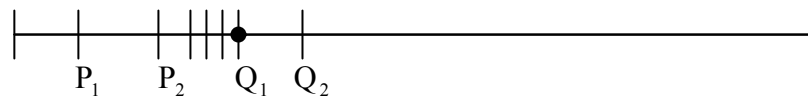
A seemingly perfectly requirement for the P cutting points is that each should have a left and right next one or the end points of I_0 . But this doesn't mean that all $Q \in I_0$ will have next cutters to its left and right. Indeed, cutting points can approach a Q .



A second problem is that even for the perfectly formed intervals between the P cutters, we have two end points, which are shared. So where should these end points belong?

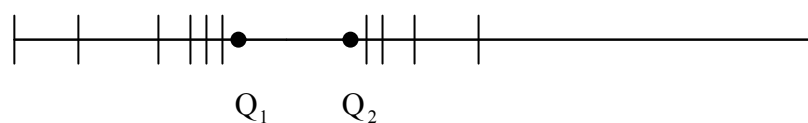
Should we agree to always include only one end or just distribute them by will?

This end point problem touches upon the previous, indeed, the seemingly missing Q can be then still the end point of a new interval:



Of course, if cutters approach Q from both left and right, then it has to be missing from I_0 .

The worst scenario is if two Q_1, Q_2 are approached from left and right, and thus, the whole $[Q_1, Q_2]$ interval is wasted from I_0 .



We could agree to only allow cutters that don't produce missing intervals. Or agree that we don't allow even missing points. Then it is a very interesting question, whether missing points only can add up to missing intervals, when rearranged.

But all these are irrelevant to our original idea. Indeed, if we cut I_0 badly and the pieces cover an S set, then it is even more true that S is smaller than I_0 . So, even more $I - S$ will be bigger than $I - I_0$, so must be a continuum.

The real problem of our idea is that it can't give new continuums!

Indeed, suppose a piece of I_0 is used. If to the left or right, there is no new piece used, then that means that S has no point there, so then $I - S$ is obviously continuum. Then if we have new pieces on both sides to any piece of I_0 , we again end up with dual choices, so an obvious continuum by our earlier principle.

In spite of this, the length argument would be an alternative and simpler way to explain the continuumness, if we could place it in a full picture. Before we turn to this, let's check again the S twins. Their limit set is the $I - S$ set if S is the cut out open windows. We already used this argument to show that non wide twins are not zero sets. Now the same argument is for the continuumness: If the total of the windows is say $I_0 = \frac{I}{2}$, then $I - S$ is at least $\frac{I}{2}$, and so continuum. For the wide twins, where $I_0 = I$, then $I - S$ is at least $I - I = 0$, which helps nothing in establishing that they are a continuum. Even though, it followed from the choice principle. So, the length principle cannot even grasp all the cases of the choice ones.

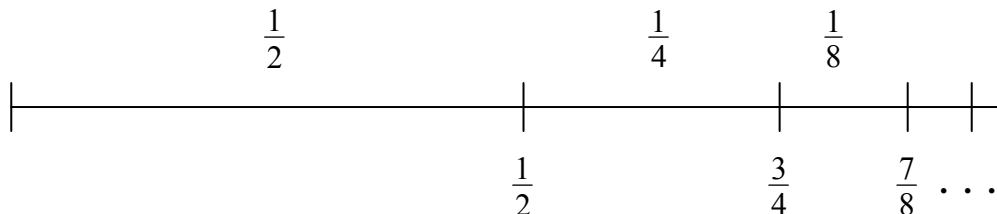
53. Nil sets

Before we go to the proof that lengths are indeed a reliable measure of point sets, we have to deal with a special case. The zero sets were coverable with finite many, arbitrary small intervals. Now that we went into the question of covering sets in an I interval with its I_0 sub-interval, the natural idea is to connect the two questions. In other words, to regard sets that are coverable with not necessarily finite many pieces, but arbitrary small in total.

Such sets will be called nil sets as opposed to zero sets.

Of course all zero sets are nil but not in reverse. The infinite many pieces can cover sets that can't be by using only finite many. But the difference is a very complicated affair! First of all we have to see that while at zero sets the arbitrary smallness was just mentioned loosely, that is not emphasizing the total, here we can't do that. For one fix finite many pieces if they are all arbitrary small then the total is too. Of course, with increasing the number of pieces it seems a bit more complicated. But not really. Indeed, if we increase the pieces from n to N , then we can require the pieces to decrease more than proportionally, so the total will be decreasing automatically too. With infinite many pieces, it's not enough to talk about arbitrary small pieces. The total must be arbitrary small too. Indeed we can use arbitrary small infinite many intervals that have an infinite total, namely using same lengths. Nevertheless the Achilles paradox shows that at least we can use arbitrary small total with infinite many pieces too.

Indeed, the simplest version of the paradox is that : $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$



If the full interval is not 1, rather an arbitrary small I_0 then the same is true:

$$\frac{I_0}{2} + \frac{I_0}{4} + \frac{I_0}{8} + \dots = I_0$$

This proportionality of the Euclidean space is one of our most natural intuitions.

The most obvious case of nil sets that are not zero is a direct consequence of this Achilles paradox with a twist, that is applied to arbitrary small I_0 .

Indeed, the natural numbers, that is points going to infinity can be at once covered.

So, unlike zero sets, which have to be bounded, nil sets don't.

The difference between bounded nil and zero sets leads to the most amazing paradoxes.

All we have to do is combine the potential of using infinite many pieces for a nil set with the simple observation that the halfers are sequencable. We went further and showed that any particular n-dividers and all n-dividers together are sequencable too.

This will make the paradox even stronger, though not visually, rather verbally.

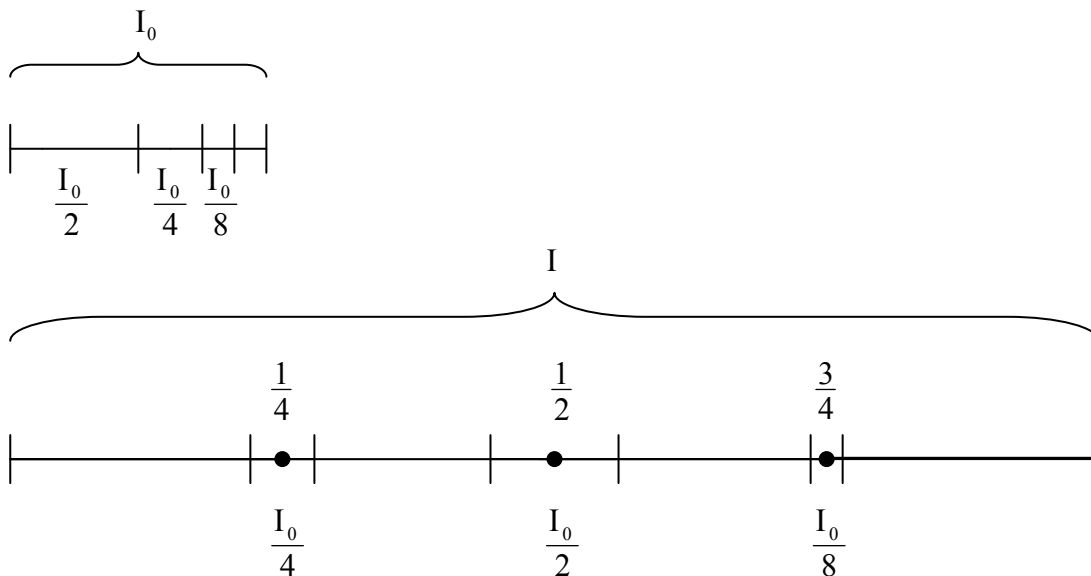
Indeed, the point is that already the halfers are a sequence that is dense in any I interval.

Lets place the consecutive smaller and smaller halves of I_0 to cover each of the halving points of I in its obvious sequencing from left to right with deepening halvings:

$$\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \frac{3}{16}, \dots$$

$$\frac{I_0}{2}, \frac{I_0}{4}, \frac{I_0}{8}, \frac{I_0}{16}, \frac{I_0}{32}, \frac{I_0}{64}, \frac{I_0}{128}, \frac{I_0}{256}, \frac{I_0}{512}, \dots$$

We can even restrict that these intervals should contain the halfers exactly at the center.



Even if we don't use any of the end points of the small intervals (the lines in the picture), the halfers will still be covered. And the total of the covering intervals is I_0 .

Since the halfers become dense in I , we have the false impression that the covering little intervals overlap perfectly and thus, cover the whole I . This of course, would mean that the arbitrarily chosen I_0 covered the full I . So then the whole length conservation that we were aiming for is hopeless. Of course, as I said, this is a false impression!

The length conservation, even though it's a less visually reliable assumption, is true!

The mistake was the assumption "overlap perfectly". There is no question about that the little intervals must overlap somewhere. For example, there will be halfers that will go inside the $\frac{I_0}{2}$

covering interval of $\frac{1}{2}$. The covers of all these will be overlapping with $\frac{I_0}{2}$. And yet, these overlappings are still negligible, compared to the directly non visible, non overlapping ones.

So, amazingly most of the points will not be covered at all.

Formally, the paradox can easily be resolved by saying that:

Even though the halfers are covered, their limits are not necessarily.

But where are the limits of the halfers anyway? Well that was an original paradox, or rather merely puzzle, that now became magnified into a full blown paradox.

In spite of this, the formal resolution by mentioning the limit points is useful if we continue it a bit even more. Why didn't the covering intervals cover the limits too? That's easy, because only a closed set contains all limits. So using the open intervals, we right away made the covering system not closed. At zero sets, the total opposite was used. We allowed closed pieces and so they covered the limits too. But, it was not only the closed pieces that we chose, that made this possible. Indeed, the vital part is that finite many closed intervals or for that matter, finite many, any closed sets together are closed, but infinite many don't have to be. So here with our infinite many pieces of I_0 , also it's not the choice that we make for the pieces to be open or closed that matters. If we would use all closed pieces, the total would be still just I_0 and couldn't cover I . The infinite many closed little intervals, exactly due to the strange partial overlappings and non overlappings, add up to a total set, that is not closed and thus, indeed will not contain the limits. On the other hand, using only open intervals, their total is an open set! Indeed, this is quite universal. All union of any open sets, is open, simply because if a point was inner in one set, it is automatically inner in the union. This increases the paradox even more: Not only all the halfers can be covered by arbitrary small total, but the total can be an open set. We can push this, even further by noticing that the complement of an open set is closed. So then, the uncovered part of I is a closed set.

The whole covering of the halfers of course only depended on their sequencing. So, any other dividers or even all dividers, that is fractions, can be covered the same way with arbitrary small open set. And then, the uncovered closed part will contain only irrationals.

So we obtained the same result that we already showed with the twins in section 27.

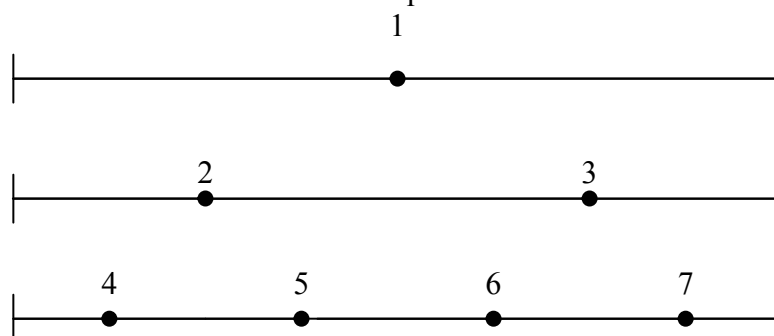
There, we also used the sequencing of the dividers to position the twins, so that all dividers were jammed into the windows between the twins. Thus, the twins and their limits had to be outside the dividers or fractions and since the twins and their limits are closed, they gave a closed subset of the irrationals. The twins gave the extra advantage of choices, thus continuumness. Here we don't care to show how twins can be inserted outside the dividers.

Rather, just by covering the dividers, we obtained a closed complement among the irrationals.

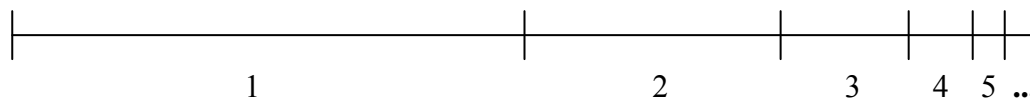
54. The didactical angle

There is a paradox that is even deeper than all the ones about points above! This is the reason that I delayed the concept of the nil set to this late point. I was struggling a lot about this decision. In a sense, it would have been much more theatrical to start with the sequencing of the halfers combined with the Achilles paradox to create the amazing arbitrary small covering of a dense set. I rather chose the twins as a round about way to approach the same. But, the fact then still remained that this shorter approach does exist. Then, the real question is, how could this avoid the attention of all the great mathematicians. So now I will make a new attempt to show that indeed, a hidden spell of veil does exist and stops even the smartest people to see the simplest problems.

Lets start from scratch. Poor Gauss will be my guinea pig again. Smartest monkey in the lab, we gotta stick with him. I want to explain that the puzzle is not an amorph claim that Gauss missed Sets or Equivalence and so on. No! Gauss never saw that the half dividers can be covered by an arbitrary small interval. To see that fact, one has to combine two, in themselves, very banal facts. One is that the halvers are a sequence:



The other is that any I_0 interval can be split into a sequence of pieces:



Both of these images can float in someone's mind and cause nothing.

The two images colliding is the click! Relating two seemingly obvious and useless makes the spark. But it is still just a spark. Yet once the spark is in, once the vision, the idea is born, that the halfers can be covered by any small I_0 , it will start the engine. And that's where the difference is between a moron and Gauss. If Gauss had seen that the halfers can be covered, then at once, he would have seen a lot of other things too. Namely, that all other dividers and all dividers together are merely a sequence too, and so can be covered the same way. Thus, the existence of irrational points would follow, avoiding all the struggles of the Greek constructions. This reaction is unavoidable. But for Gauss, who was one of the smartest brains ever, the reactions would have had to go much further. He would have realized at once, that the seemingly obvious assumption that the full interval cannot be covered this way, hides the wider question of why an interval cannot be covered with pieces of a smaller at all. All this is immediate. Going through common point axioms, using choice sequences, are elaborate ways. The spark is immediate. If you are smart enough to see the consequences of a spark then the new axioms and theories can be formed in minutes. Gauss was obviously smart enough to see the consequences further than anybody else. He would have invented equivalence, the continuum, measure, all in a flash. The flow is unstoppable. And yet the spark is everything. And it can be missed, by simply not seeing something obvious. This is a veil. This is a spell. This is the emperor's clothes. Now this was only a philosophy about the act of creation.

The didactical angle is much more complicated. To break the shield of mediocracy, is harder than enjoying particular sparks fly. So to raise the question of nil cover as the best way to open the lid of equivalence and cover or length invariance is unquestionable. You can't jump through this and understand what's going on. It's a didactical block, a fixpoint of understanding. And yet Formalism does the opposite! Instead of bringing it into the light, it hides it. Why? Very simple. Because once the equivalence and the rest are introduced, it wants to go as "smoothly" as possible. As convincingly by the logic as possible, and caring not about the vision at all. So why bother to provoke the complicatedness of this whole beehive, whether length can become bigger or not. This will just be smoothly melted into measure and rest in peace.

A dialectical fact is this: The shortest way to prove that a full interval is not sequencable becomes the longest, both formally and didactically. Formally because the depth of measure lies in it, as an unborn child. Didactically because all concepts of point sets are hidden in it too. So here is this strangely perfect "best and worst" start to point sets:

An I interval cannot be sequenced as P_1, P_2, P_3, \dots because then we could cover

P_1 with $\frac{I_0}{2}$, P_2 with $\frac{I_0}{4}$, P_3 with $\frac{I_0}{8}$, \dots , P_n with $\frac{I_0}{2^n}$, \dots and so on.

Then, $\frac{I_0}{2} + \frac{I_0}{4} + \frac{I_0}{8} + \dots = I_0$ would cover I , regardless of the size of I_0 .

Indeed, it provokes everything! Achilles paradox, equivalences, density, continuity, measures.

As I mentioned, I was contemplating to start the point sets this way. Doing that, still wouldn't have made me a Formalist in itself. Yes, the proof hides everything, most deeply the problems of density. But I could have approached density from these different directions. I found it too messy, that's why I chose to start with density of a set in its three clear cut versions as:

at a point, in itself, and in an interval. But you can paint a same picture in different ways!

The short simple proof above as a logical process did not necessitate the equivalences and densities! All we claimed is that I is not P_1, P_2, P_3, \dots . Where and how such sequence can be in I , is immaterial. The only logically missing part is the assumption that to cover I with an arbitrary I_0 is impossible.

The Formalist travels through these logical necessities as a pretence to hide certain visions. The non formalist travels in visions he wants to share! But the final verdict that makes the Formalist false while the non formalist true is the following: Seeing the right pictures, everybody can create the right proofs without the false tools of rememberings and hidings. But the reverse is false. The correct logic contains only traces of the pictures. For two different reasons: One is that the logic of proofs can hide even the pictures of what we prove. And secondly, we always have unsolved, unproved problems, that are very much involved in the pictures. It is true that some Formalists do like to mention the unsolved problems, but only as logical holes. In fact the difficulty is their only concern. So these people will never mention easy facts that paint a picture. The most important distinction that separates the Formalist is the intention. The original mathematicians, who created new visions, had to use final forms to express the exactness of their new common sense. The epigone textbook writing morons, who are in the sick abstract tertiary competition of being hyper precise, don't discover anything new. Their books roll off the conveyor belt every month with the same regurgitated content. But nowhere is so evident this phony exactness as in Topology, the mathematics of point sets. This should be the most visual. Instead it became the most abstract, overdoing algebra. The end of the line is the Wikipedia math articles about point sets. They don't have the excuse not to be educational. They simply don't care! They lost the plot! Here is an example of insanity out of control: Look up the Brouwer Fixpoint Theorem! After the usual visual explanation about the meaning with maps crumpled the following proof is given:

Assume that there does exist a map from $f : B^n \rightarrow B^n$ with no fixed point. Then let $g(x)$ be the following map: Start at $f(x)$, draw the ray going through x and then let $g(x)$ be the first intersection of that line with the sphere. This map is continuous and well defined only because f fixes no point. Also, it is not hard to see that it must be the identity on the boundary sphere. Thus we have a map $g : B^n \rightarrow S^{n-1}$, which is the identity on $S^{n-1} = \partial B^n$, that is, a retraction. Now, if $i : S^{n-1} \rightarrow B^n$ is the inclusion map, $g \circ i = \text{id}_{S^{n-1}}$. Applying the reduced homology functor, we find that $g_* \circ i_* = \text{id}_{\tilde{H}_{n-1}(S^{n-1})}$, where $*$ indicates the induced map on homology.

But, it is a well-known fact that $\tilde{H}_{n-1}(B^n) = 0$ (since B^n is contractible), and that $\tilde{H}_{n-1}(S^{n-1}) = \mathbb{Z}$. Thus we have an isomorphism of a non-zero group onto itself factoring through a trivial group, which is clearly impossible. Thus we have a contradiction, and no such map f exists.

Being didactical is very hard! It is not appreciated in this phase of mankind. In fact it is intentionally avoided. In a selfish world why should knowledge be given away. The petty low life Formalism of the mediocre educational Nazis has nothing to do with the sins of the geniuses. These words were well chosen!

Fascism was halfly defined by Dimitrov! The petty bourgeoisie did not go away as a category just because the second world war ended. The trick of Hitler was absorbed by the winners. The seemingly opposite democratic freedom, actually hides the same essence. A new mindless and scared class, the neo-proletar. Striving for just enough privilege, that ensures the comfortable advantage of consuming things without creating things. Foods, clothes, cars, thoughts.

The other half, that Dimitrov missed, is the economic nationalism. The war-economy that is hiding behind the seemingly open globalization, just as a depraved paralyzed mass is hiding behind the democratic freedom. It's hard to believe that at the root of all this, lies the main category of Formalism. So lets return to mathematical Formalism.

This would only continue our simple proof above, with the logical gap. Why an arbitrary I_0 shouldn't cover an I . We talked about all the hidden visions, so we have to face this too.

By the way, there could be three possible reactions to this missing part:

Some would simply accept it as trivial, that a smaller I_0 cant be cut to cover a bigger I .

Some would say, that cutting into pieces is messy enough to make the increase of length possible. Finally some could say that, even if a fix smaller I_0 could cover I , to do this for arbitrary small I_0 is impossible. This last is the least mathematical minded view.

Indeed, if a smaller I_0 can cover an I , then we can use an even smaller to cover I_0 . Then again, and again, so it would be plausible to use arbitrary small cover too. Its not quite accurate argument, because a length can decrease without going to zero. But the conclusion is correct: So the truth is that any length increase is impossible but not evidently!

We have to and we can prove it!

55. Open cover

The heart of the idea is a new continuation of the already known complementarity.

At the zero sets, as I mentioned before, the closedness of the cover was guaranteed, only because we used finite many pieces. On the other hand, the total of even infinite many open sets is open. So here comes a new complementarity: For covers that allow infinite many pieces, we should use open intervals. This of course, will not guarantee the covering of the limits as a closed cover would, so we have to find a new advantage for an open cover.

It is very simple and surprising, in fact, in a strange way, resembles the zero cover.

T

Borel Covering Theorem

- 1.) If a closed I interval or cube in space is covered by any set of open sets, then already finite many of them will cover.

Amazingly, the grand scheme of complementarity comes through in an even further step, because the theorem at once implies a generalization:

- 2.) If a closed S set in an interval or cube, is covered by open sets, then again, finite many is enough to cover S .

P

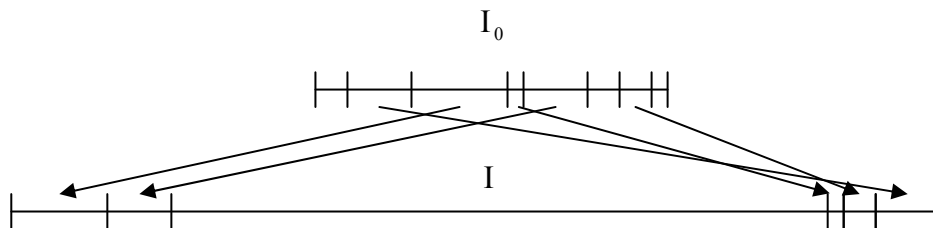
The actual proof of 1.) will be in the next sections.

The mentioned resemblance to zero sets is not just the enoughness of finite many covering sets. To see the deeper resemblance, we'll create a new consequence of the Achilles paradox and show why the I interval had to be closed. We'll cover the $[0, 1)$ unit interval without its end at 1, with a system of open sets, from which definitely finite many can't be enough.

The open sets will be all the same long open intervals merely shifted forward, again and again, with the consecutive halvings of $\frac{1}{2}$, that is with $\frac{1}{4}, \frac{1}{8}, \dots$. In order to cover 0, we use as first interval any $(-x, \frac{1}{2})$. So then the next will be $(-x + \frac{1}{4}, \frac{1}{2} + \frac{1}{4}) = (-x + \frac{1}{4}, \frac{3}{4})$ then $(-x + \frac{1}{4} + \frac{1}{8}, \frac{3}{4} + \frac{1}{8}) = (-x + \frac{3}{8}, \frac{7}{8})$ and so on. They clearly cover $[0, 1)$, but finite many of them couldn't go "all the way" towards 1. Yet, amazingly if 1 had been included, that is the closed $[0, 1]$ had been covered by a system, then the covering of 1 would cover the approaching ends towards 1 too from a point, so then finite many would indeed be enough.

Now, lets see why 2.) is immediate from 1.).

If an S closed set is left out, the remaining $I - S$ will be open. Now a cover of S , can easily be made into a cover of I by simply adopting $I - S$ as a new member of the cover set. Then if finite many members cover the whole I , we simply leave out this $I - S$, if it were among the finite. Clearly, leaving this out, will not affect the covering of S by the remaining ones, because this $I - S$ only covered points outside S . This simple generalization also shows that the essence is the first claim, the finite coverability of an I . Also, the first form is directly relating to our original problem of why an interval shouldn't be coverable by pieces of a smaller I_0 . In fact, we can show at once that it must be impossible. Suppose I_0 could cover I :



Lets not worry about how the end points are distributed. Similarly, in space, how the surfaces would be distributed. Lets be generous! Increase I_0 a bit to an I_0^+ that is still less than I .

In fact, use the same pieces in I_0^+ as in I_0 except increase all a tiny bit in both ends, or in space in every direction. These will obviously cover the older pieces of I_0 .

Now, let's leave out the end points or surfaces in the I_0^+ system, to obtain I_0^{+-} .

The pieces of I_0^{+-} are all open intervals and still will cover all the pieces of I_0 .

Thus, we obtained an open cover system, that covers all pieces of I_0 and therefore I too.

We can apply 1.) and so, finite many should cover I . But this is a sheer contradiction in the addition of lengths or volumes. Indeed, it boils down to two members:

If $a = a'$, $b = b'$, $a + b \leq I_0^{+-}$, $a' + b' \geq I$ then,

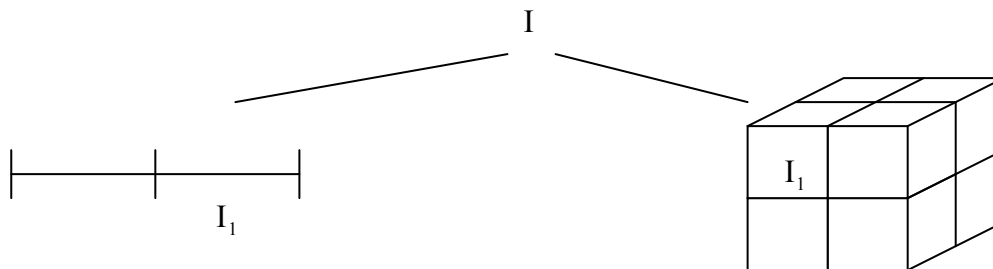
$I_0^{+-} \geq a + b = a' + b' \geq I$ so $I_0^{+-} \geq I$ contradicting our assumptions.

56. The short proof, continuing against Formalism

Here is the formal proof for 1.). It is short, striking and stupid. Total indirectness.

Suppose no finite subset of the covering system would cover I .

Cut I into two halves or eight quadrants in space:



Now if there were finite covering subsets for the two halves or for the eight quadrants, then combining these would be also a finite subset that covers the whole I . This means that with our assumption, that no finite covering subset exists for I , at least one of the halves or one of the eight quadrants, I_1 , again cannot be covered by finite subset. Cutting I_1 again, we can repeat the argument and the non coverability by finite subset, inherits again to an I_2 , and so on.

These narrowing I_1, I_2, \dots intervals contain a final P point. This P must be covered by an S member of the cover system. Since S is open, P must be inner point of S . Thus, S must not only contain P , but all the narrowing intervals from a small enough size: I_n, I_{n+1}, \dots

But these were all uncoverable by a finite subset of the system and now the single S covers them all. Total contradiction. So the assumption of no finite cover of I , had to be false already.

I intentionally used even in this formal proof, the expression "total" contradiction. A real Formalist would have merely said, "There has to be a small enough I_n inside S . Thus, S is a finite cover, contradicting that I_n doesn't have one." He wouldn't emphasize that all the

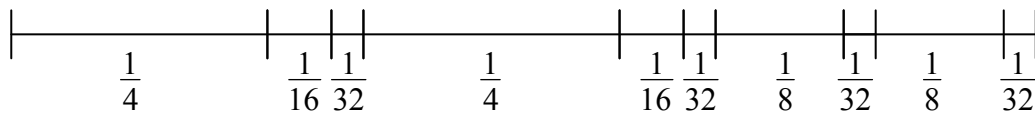
I_{n+1}, I_{n+2}, \dots are giving the same contradiction, and that single cover came about.

I am not against this proof. Formalism has its role. Here I used capital F , because it started the sentence. But it was merely formalism. It is Formalism that has to be eliminated from our world gone astray. The difference is simple. You have to be honest and truthful. It's not what you say that makes you a Formalist, it's what you don't. So there is nothing wrong with the above proof if you continue. Dig deeper. But most importantly, the false belief that you can't or shouldn't dig deeper is the evil. The stubborn, but sly anti philosophical puritanism. Some assumed common sense, covering the incapacity or laziness to dig deeper. Or a utilitarian justification, that it is better not to over explain, that a struggle of understanding is healthy. This is the dangerous crap that is infectious, like the turning of victims into new bullies, by the military institutions. This is the poison that makes someone say in memory of Shoenfield that the "examples of this book made me into a mathematician". Well good for you! Shoenfield in the preface of his book, said thanks to some people who opened his eyes to the fact "... that Mathematical Logic is not a collection of vaguely related results, but a method of attacking some of the most interesting problems that face the mathematician." I'm sure those letters and conversations were very eye opening. But did he intend to open other people's eyes? No! He

wanted to jam as much result as possible into his “big book”. And he did. His is the best Logic book. Best of the worst, because there are no good ones. I can spit acid and swear against these false educators without end, but the bottom line is simple. You shouldn’t lie! Shoenfield knew from the first letter of his book, that the chosen axioms, the so called variable logic is a mere consequence of proof theory. That there is a bigger picture of “ands” and “ors”, that give trivial completeness and from which the implicative logic follows as particular. When Hilbert discovered his rules of quantors, he had no choice. He had a hunch, he was a formalist, just like Euclid when he stated his parallelity axiom. The same Hilbert said that mathematical truth can only be something you can explain to the first person you see on the street. And he meant it. Even Kleene’s classic book that Shoenfield praises was attemptive to teach. Though, he went astray by tying up the loose ends, instead of loosening them even more. It’s very easy to establish how much someone sees, and how much he communicates from that. This is what determines if someone is a mere formalist or a Formalist. This category then grows out of mathematics, and indeed is the most fundamental at present. The bureaucrats, the opportunists, the petty bourgeoisie, are all Formalists. The economic version of Formalism is the privilege or advantage, using what one has or knows to increase his extra status. The extra profit ratio of Marx, is the simplest consequence. The fact that he falsely thought it merely originates from the owners of production is immaterial. Today, its obvious, that the Beast is independent of people. The grannies investing in shares are just as capitalists as the Rockefellers were in Marx’s time. The stock market is the Beast that creates the false values as the prices. But Marx knew that too. His fight for the existence of value is what made him a true Idealist, in spite of his self assessed materialism. Socialism was not his idealism, but merely a mistake in his idealism. In spite of this, Socialism as a temporary historical reality, was a realization that value can be defined. The prices of every single spoon and fork in the stainless steel, in rubel and kopek, staying forever, were and are the brave defiance of sanity against greed. Of course, it didn’t work. The liars and buyers, they are everywhere. The bad has to be worse, till the emperor’s clothes falls. But the bottom line is simple. A lie is a lie. How we locate the lie and despise the liar is a difficult task. Newton was never lying in his Principia. He struggled so hard to explain. Did he fail? Yes. He could have said: Newton Laws, 1, 2, 3 plus gravitation, writing a bad high school physics book. But he knew there was more. There is space, there is energy. Apart from all this, he was a bad person. So was he a Formalist? Yes. In the bigger picture. Gauss quite contrary, was not a bad person. He was a Formalist as a mathematician. In the bigger scheme of things, they were both just Formalists. Failed miserably and became idols of success. The morsels of truth survive in this false historical picture. Pros and cons, little details. The big truth that every individual has to recreate creation, discover the lies and hidden angles, is the only taboo. Not the fact of lies. That is open. That is the norm of maturation. The essence of all the years spent in schools. Puberty, sex, is the biological seal of approval. The ultimate game of lies. In that, the privileges themselves become an organic unity of biology and society. The looks, the status, the bigger better deals, the safety of family. And most of all, no questions about the obvious contradictions in them. But this is not taboo, it’s just a denial of consciousness. This is the obvious that makes a TV show funny or sad. Of course, it’s hard to “show” it without spelling it out. So indeed, it’s not the subject, it’s not the what, but the how that matters. The show about nothing. The taboo is real, it’s not the fact of lies, it’s the content of lies. That every single truth is sealed in a bottle. Only human interaction, teaching beyond telling some truth, revealing and giving away our privilege can do that. The biggest lie is that by merely sharing things or money, we can still be honest. Every attempt towards equality, that doesn’t involve the truth of human conditions, is only increasing the inequality. This at once means that social justice is a contradiction. Which of course is obvious in all of its failing attempts. All points of social contacts are pretence. Even whole institutions are false. Schools don’t bring out the best from the children, they put in the worst. Prisons don’t change the bad in people, they punish them for it. The utilitarian middle of course gives way to comedies and dramas. Schools just teach you to get better jobs. Prisons just keep away the criminals for a while.

57. New proofs, refinements

Lets call a refinement of I any cutting it into sub-intervals, that are all halves, quarters, or result of higher halvings.



In space similarly, we can combine a cube from repeated eight cutted pieces.

We can use arbitrary small pieces, which of course means that such patchworking refinement of I can contain infinite many pieces. A refinement is finite, if it only contains finite many pieces. In this case, there is a smallest size of patches in it, though more such minimal patch can be in it. In fact, we can have a uniform refinement with all same sized pieces.

Obviously, every finite refinement can be further refined to make such uniform refinement.

A refinement is individually covered by the open cover system of I , if every patch in it is covered by a single open set. Obviously if a refinement is individually covered, then at once we get a cover of I , by picking for each patch in the refinement, an individual covering open set.

In particular, if a finite refinement is individually covered, then this gives a finite cover of I .

So it's enough to show that I has a finite, individually covered refinement. Or due to an also obvious uniformization that: I has a uniform, individually covered refinement.

The opposite of this would be that for every uniform patching, at least one patch is not covered by a single open set. Now we show that this is impossible.

The existence of such uncovered patch, is inheriting again from I to at least one of its halves.

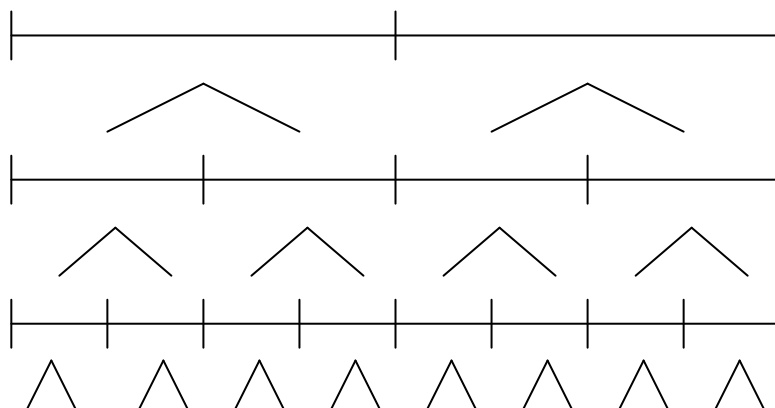
Indeed, obviously if both halves had perfect uniform refinement, without such uncovered patch, then we could combine them, using the size of the smaller ones. So, repeatedly, we obtain a narrowing halvings that all don't have uniform, individually covered refinements.

But these narrow to a P point, which is in a single open set, and so a narrowing halving is also inside. That then is individually covered, contradicting that it wasn't.

So what was the point of this new proof? Just as indirect, but much longer.

Actually, we gained something! We saw that an individually covered, finite refinement is the cause of why I can be covered finitely. So now, we should see too, why such finite refinement must exist, or even better, how we can obtain one.

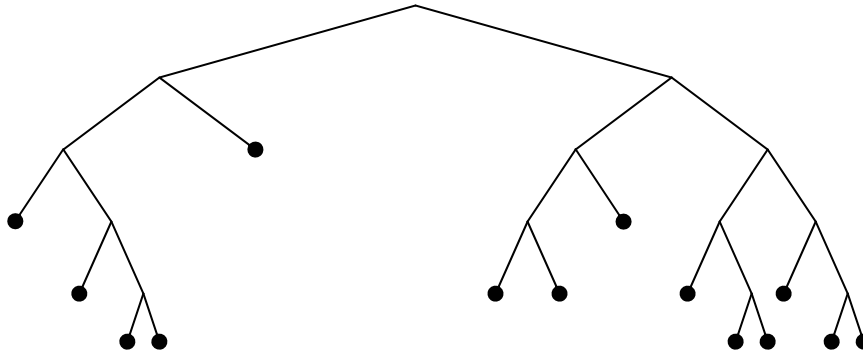
Lets regard the potential refinements of I simultaneously:



Once a patch enters into an individual open cover, we should cut the branchings off.

Can we still have an infinite narrowing? No, because it goes to a P point that is covered and so, before P is covered, already a patch is covered too. So we don't have infinite path in the branchings. Does that mean, that we have a maximal length in our branchings?

If that's the case, then we are finished, because then we only have finite many ends that give exactly the patches of an individually covered refinement.



So amazingly, our refinement follows from a quite independent and universal fact:

A dual branching system without infinite path, must have a longest maximal path. In short:

A dual branching system without infinite path is merely finite.

The negative form of course is that if we have longer and longer paths, that is the branching system is infinite, then we must have an actual infinite path too.

Observe that having infinite path is not visually obvious from the system being infinite.

Indeed, we could very well imagine that longer and longer paths exist, without a single infinite.

An even better way to see the non obviousness of the claim, is imagining that we actually make decisions at every branching which way to go. If we make a bad decision and go to a dead end, then we have to go back and reconsider every earlier decision we made. The infiniteness only guarantees that going back and changing some bad decisions, we can go longer than before.

But this doesn't imply at all a perfect sequence of decisions that leads to an infinite path.

If at every branching, a little fairy could help us by answering one question, then what should we ask? Is there a strategy to guarantee the infinite path without returns? Of course, we could ask: "Which way is the infinite path?". But I don't think the fairies would allow that. Instead, we can ask this: "Which way is an arbitrary long path?". This is not an unreasonable question to ask, because we were given the fact that there are arbitrary long paths right from the beginning.

So at the first branching, one definitely must continue with arbitrary long paths. Then in that choice, the next branching again has to have at least one, that still allows arbitrary long ones, since we were promised one before, and we only went one step ahead. And so on, the arbitrary long continuation is a possibility, that must inherit to always at least one of the choices. And voila, it also guarantees no dead end. We got a strategy!

So we couldn't get a positive proof, a definite form of the individually covered refinement. But we realized that somehow the choices come in again. Thus, the earlier totally indirect proof, now had a strategy of refutation at least. In fact, with these hindsight, we can fabricate a positive proof, just as short as the first. All we have to do is reformulate our claim:

If a system of open sets doesn't have finite sub cover of I , then there is $P \in I$ uncovered.

Such P can be given quite explicitly: Always choose a halving of I , in which we don't have finite sub cover. These halvings are the path that leads to P .

58. Rare sum

Now that we showed that an I can't be covered by a smaller I_0 , it implies that no I can be nil set. Since every P_1, P_2, \dots sequence is nil set, thus an I is unsequencable.

An unfinished business is the relation of nil sets to zero sets.

The dividers are clearly nil set, but not zero. They are dense.

Can nil sets be not dense, or to go all the way rare? Of course! We saw the wide twins, they were zero, thus nil too. So the real question is, can a rare set be nil but not zero. We saw that too. They are the twin splitters. They are still just a sequence, so they are nil. All these examples were sequences. But neither nil nor zero doesn't have to be sequencable. Indeed, the zero sequences keep their limits, so the wide twins with their limits are zero and thus, nil continuums too. Since a subset of a zero set is zero too, thus we can't say that among zero sets, all sets are continuum or sequence, like among the closed. Still, in a sense, the zero sets are not good to search for non continuums that are non sequencable, due to the conservation of the limits.

The nil sets are much better. This brings up a simpler question, we already asked. Whether all continuum nil sets have to be zero. No again. We can combine a sequence of zero continuums densely, that is placing them like halfers. The total is still a nil set, because we can repeat the $\frac{I_0}{2} + \frac{I_0}{4} + \dots$ idea. That is, split each of these into infinite many again:

$$\frac{I_0}{2} = \frac{I_0}{4} + \frac{I_0}{8} + \dots$$

$$\frac{I_0}{4} = \frac{I_0}{8} + \frac{I_0}{16} + \dots$$

Thus, a good idea for a new non sequencable set, that is a candidate to refute the continuum hypothesis, would be a nil set that is not even a combined sequence of zero sets. The trick to do this is again, aim higher, namely get one that is not even a combined sequence of any rare sets.

Indeed, remember that all zero sets are rare!

Lets call a combined sequence of rare sets, a “rare sum”.

Rare sums are an ingenious big generalization of sequences!

Indeed, a single point is an obvious rare set, so sequences are all rare sums.

Amazingly, first of all:

The concept of rare sums yields an instant generalization of the non sequencability of intervals:

T Rare sum cannot contain a full I .

P Suppose $I \subseteq S_1 \cup S_2 \cup \dots$

S_1 is nowhere dense, so in I too, so there is an I_1 in I , outside S_1 .

S_2 is nowhere dense, so in I_1 too, so there is an I_2 in I_1 , outside S_2 .

And so on, we get an $I \supseteq I_1 \supseteq I_2 \supseteq \dots$

A common P point of these is outside all the S_1, S_2, \dots

T Baire

If $S \subseteq I$ is rare sum, then $I - S$ is not.

P $S = S_1 \cup S_2 \cup \dots$

Suppose $I - S = T_1 \cup T_2 \cup \dots$ were similarly.

Then, $I = S_1 \cup T_1 \cup S_2 \cup T_2 \cup \dots$ would contradict the above theorem.

The reverse of this theorem is obviously not true. Indeed, half of an interval is not rare sum, and the other half is neither. The simplest application of the theorem is that the irrationals are not a rare sum. Indeed, the rationals are obviously, because they are a sequence.

Here the rare sum is a nil set, so the irrationals are 1-sized non rare sum in $[0, 1]$.

To get our promised weird example, we reverse this situation.

The rare sum will be 1-sized, so the complement becomes a non rare sum nil set.

To get 1-sized rare sum, we combine a sequence of rare sets that are $\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \dots$ sized.

The first is the leftover set of $[0, 1]$ with $\frac{1}{2}$ total windows carved out between twins.

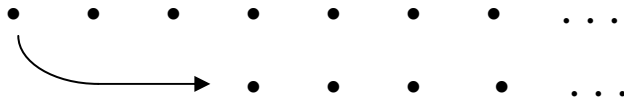
The second is the leftover, with $\frac{1}{4}$ total windows, the third with $\frac{1}{8}$ total windows, and so on.

Thus, our nil set, that is non rare sum is directly visible too, as the common part of the open window sets. Observe, that this total, not only isn't an open set, it cant contain inner point at all. So, this set cannot be trivially continuum, but it can still be if, if it contained an infinite sequence of choices for points. Unfortunately, we can't prove that it is not the case.

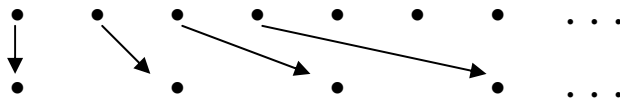
59. Lost points: Jigsaw paradox

Now we return to the simple question, “can we lose points?”. This is very appropriate. We demonstrated many weird point sets, but they are all hard to visualize point by point. Of course, this question of losing points, is meaningless without telling where we lose the points from, and by what changes.

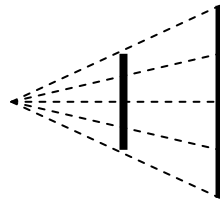
We know for example, that an infinite set can lose elements by shifting it.



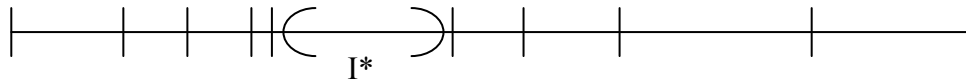
By combing it into itself:



We even saw the specific point loss of continuous projections:



When we started the interval cuttings, we mentioned that cutters can approach an I^* interval from both ends:



This I^* open sub interval is not determined by cutters as end points.

This was unimportant then, because we only wanted to cover some S sets, so not using I^* still left our main argument intact. Now that we pursue the question of losing points, we want to use every interval that comes about by the cutters.

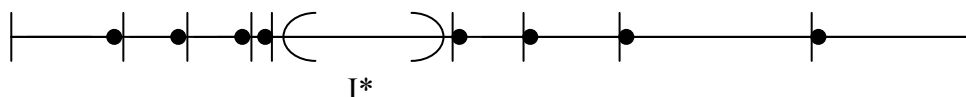
So the question is whether reassembling all the intervals, can we lose points?

This approach is also logical, because we established that the length cannot increase, so an I cut and reassembled, can only be a same length I with more or less points.

If there can be lost or gained points, then this should be called the “Jigsaw paradox”, because the pieces are very similar to that.

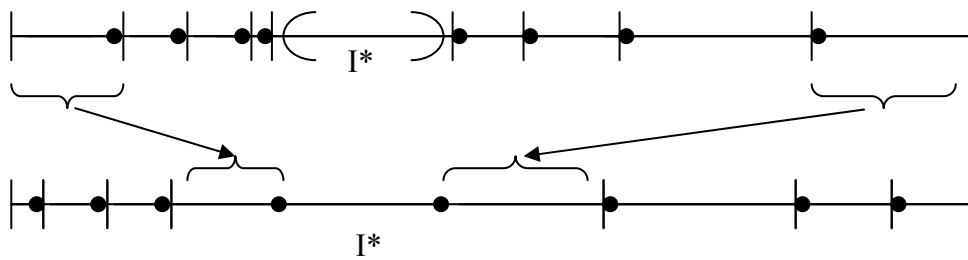
If the approached I^* subinterval above can be used to reassemble the complete I , then we lost the two end points of I^* .

The cutters themselves must be used, so we have to decide the distribution of end points in the intervals between the cutters. The trick starts with the observation, that if we decide which side the cutters should go, then we can only keep one end in all intervals, except the two at the ends of the original interval. Indeed, they don’t have next cutter anymore, so the two end points of the original I , can be kept. So keep those ends there, as if they were cut off from the two intervals at the left and right end. Instead, keep the other end points of those intervals, which were indeed cutters. Then, again keep the other end cutters of the next intervals, and so on:



The black dots in the picture show the cutter distributions. Now comes the real trick:

Reassemble the intervals in both sides, in opposite order. That is, starting the left and right end ones in the center, and approaching instead of I^* , the two kept end points of I :



The missing end points of I^* are now covered by the end cutters of the moved left and right end intervals. The rest of them go in chain, so there won't be any missing end points.

But what's more, the ends of I are approached, so the loss is swallowed up.

Thus indeed, the two holes at the end of I^* disappeared!

A simpler version of this process is using the good old repeated halvings of I with only beginning end points.



They continue each other perfectly, but I has no right end point, only left.

Now put them backwards.



Now I has no end points on both ends!

Putting such point losses into sub intervals, we can lose or gain infinite many points.

Since the cutting points must have next ones, in other words, are isolated, they can only be a sequence. So we can lose or gain only a sequence of points. This shows that this actual loss of points shouldn't be confused with the mere subjective "loss" of points, from our attention.

Cutting out windows from an I between the twins leaves a continuum:

This remaining set is the twins and their limits. These limits are not visible to our attention, but they are not lost. In fact, they are a continuum, so more than the twins themselves.

But the twins are not acceptable cutters. They don't have a next one to their left and right.

In fact, each has exactly only one to their left or to the right. In addition to this, using wide twins, that is windows, that add up to the full I size, makes the twins and their limits a zero set.

So we might say, that a continuum disappeared. But this is not, in a jigsaw re-assembling sense.

Or if you wish, you can say that the jigsaw principle is stupid, it allows point losses, so these just melt into the other subjective "point losses" with size conservation.

60. Rigid collapse, a new direction

I don't think that the jigsaw principle is stupid. Every little piece was an actual interval, allowing only the freedom of which end points belong, left or right, or both or none. Keeping them exactly with this choice, we can put them together again and lose or gain points.

If this is not a paradox, then what is? But we can go along with the criticism, that moving away all the pieces in different directions is a bit fishy. So now we introduce an opposite idea to the jigsaw principle. We allow individual points instead of intervals, which of course, would allow the absurdities of projections at once, but now we require that the points must move together.

This of course makes no sense at all, because if all points of an I move together, then we merely have a displacement of the whole I to somewhere else. So the real trick is to allow finite many "groups of points". Each group indeed must move together, but the different groups

can move in different ways. This is exactly like moving a car by moving its chassis, its under carry, its engine, the seats, and so on, and then putting it back together. No infinity, no funny business. The groups must stay together, but the different groups are moved on their own. The only generalization from a car is that we allow, in a group, any points as pieces.

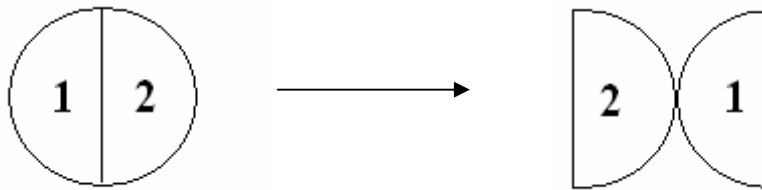
For example, the lights could be a group. So we would place them on a chart paper, exactly in the same distances to each other, as they are on the car, and then move them together. In fact, that's what we would do in real life too, except the chart paper and the lights, could be put in a little package. Then, when moved to the new location, we know where to place them.

In mathematical form, the I interval or any other continuous object, should be distributed into finite many of its subsets, and then taking these subsets apart, but keeping each exactly as they are, we could re-assemble the finite many subsets again, somewhere else.

Great! So we made a mathematical representation of household removals. What's the point?

Obviously, the subsets will re-assemble into the same full object somewhere else.

Okay, no question about this, as mere removal strategy by subsets. But is it possible that a different assembly of the subsets, can create a different object? Well of course, look at this:

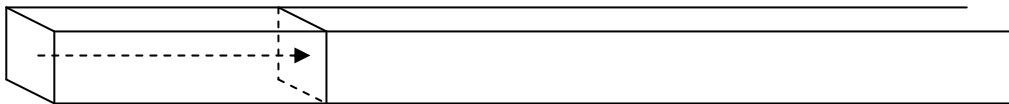


Or think of Frankenstein! So now a new question. Is it possible that the subsets can be put together, so that they become not a totally new object, rather an object that is part of the original?

Or in mathematically: Can finite many subsets be reassembled to become a real subset?

Can we lose points by rigidly moving finite many subsets?

One fairly obvious solution can easily be created, simply by the already known shifting of a sequence into itself. Using a continuous infinite long rod, can also be moved in itself:



We didn't even need the different finite many subsets, the whole set lost in itself.

Of course, an infinite rod is hardly an object, so a new question would be, whether a bounded continuous set can do the same.

Since the bounded set has left most and right most points, clearly a shift to the left or right, will not leave these inside the object. But, we could use other rigid motions, and most importantly, we could cut the objects into finite many parts, and use different motions for each.

The infinite rod case above, besides being a too easy solution, gives an additional and useful angle. The lost beginning part of the rod, was negligible compared to the infinite rest, so then an interesting question is, whether among the bounded solutions that we look for, can it be that the S object not only becomes its S_0 subset, but can become the $S - S_0$ other half too.

Using S_1, S_2, \dots, S_n subsets as groups that are moved, then such rigid "half" collapse paradox would be:

$$\begin{aligned} S_1 \cup S_2 \cup \dots \cup S_n &= S \\ S_1 \cup S_2 \cup \dots \cup S_n &= S_0 \\ S_1 \cup S_2 \cup \dots \cup S_n &= S - S_0 \end{aligned}$$

The three lines of "same" sets on the left are confusing. They are identical, but their locations in the S or S_0 or $S - S_0$ sets are different. They are moved away versions of each other.

This brings me to the question, why I called this section, "Rigid collapse".

If you ask a physicist what rigid means, he might talk about resistance against forces.

But to a mathematician, forces are immaterial. We don't care if an object resists the forces, because it is so strong or sturdy, or so ingeniously designed, that it repels all forces and so might even be light as a feather. All we care is that its points don't move relative to each other. All internal distances, remain the same.

Now if we ask the physicist about collapse, then again we might hear about breakage and so on. To us, it means merely the opposite of rigid, namely a collapse into itself, that is, all points staying within the original object, but not occupying the full.

But then if collapse is contradicting rigid, then why even bother? How can it happen together? Well that's easy! In the short name, rigid collapse, we didn't mean that a body is rigid and collapsed, but that a body is split into finite many parts, all staying rigid, yet in total, collapsed. So there is no contradiction, only to our intuitions.

61. Language barrier and identity crisis, the doubling paradox

I already mentioned way back at the fractions that $\frac{2}{3} = \frac{4}{6}$ causes nausea in some students.

I'm not joking. This is the natural reaction of a healthy mind against something disturbing, that is not resolved. This is very deep and usually it is merely jumped over or swept under the carpet. Equality, even in Mathematical Logic, is a tricky business. Then it mixes up with the Set Theoretical equality, which simply means having the same elements. Hearing this already should be disturbing, because how can equality mean the same elements, when the sameness of the elements means itself equality. So we define equality by equality? Yes, in a sense. We merely say that the equality of the elements inherits to the full set. In a more physical representation, this also means, that sets are spaceless. Putting the same elements together in different ways doesn't matter. So for example above, even though

$S_1 \cup S_2 \cup \dots \cup S_n = S$ and $S_1 \cup S_2 \cup \dots \cup S_n = S_0$, the set of the parts, that is $\{S_1, S_2, \dots, S_n\}$ is the same in both. They are only assembled differently. The identity of the two left sides hides this difference in an assembly in the unions. Indeed, the two equalities only express that the unions become S and S_0 , but doesn't tell the vital constructions.

The fact that everything in math is merely a set, simply means that we can attach new elements to the sets, that distinguish the same element sets, with different special combinations.

So when we talked about the halving points as a set, we didn't merely mean the collection of those points as dots, rather with their locations. So language was hidden as package to a vision.

The same way in the S_1, S_2, \dots, S_n , the locations of the points are hidden. They will be revealed when we give these sets. To use the same letters in:

$S_1 \cup S_2 \cup \dots \cup S_n = S$ and $S_1 \cup S_2 \cup \dots \cup S_n = S_0$ thus, is totally meaningful because the locations of the points are the same in each S_i , in both equality. Even though the full S_i are removed. But in a formal way, the two equality means that $S = S_0$, that is a set is equal to its part. Is that a contradiction? No, not more than $\frac{2}{3} = \frac{4}{6}$.

From a language view, two symbols being equal, may mean that we assumed them to be different, but they turned out to be the same. But in these cases, equality is merely a relation.

The $S_1 \cup S_2 \cup \dots \cup S_n = S$ means that the total points with their locations in

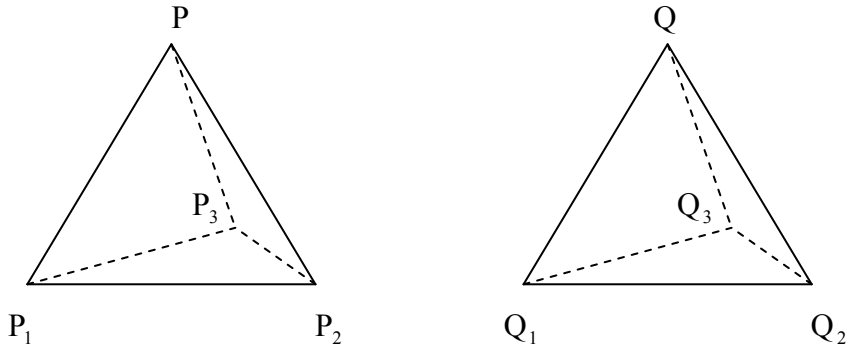
S_1, S_2, \dots, S_n are exactly all the points of S , with their new locations in S .

The points are not even important. Every location is a point anyway. The ordering of the inside locations of S_1, S_2, \dots, S_n , to locations in S , is the proof of the assembling.

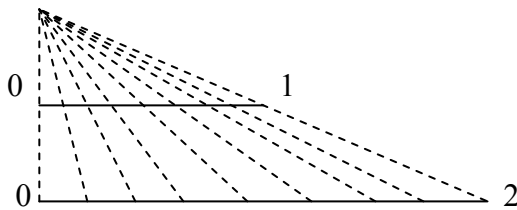
Through $S_1 \cup S_2 \cup \dots \cup S_n = S_0$, we can order the inside locations of S , to only give locations in S_0 . So, $S = S_0$ is exactly true, if $=$ is meant by relocations.

The mere fact that S 's locations can be relocated to S_0 's is not even exciting.

The projections can do that for any two intervals. So maybe the whole reassembling of S to S_0 is not a big deal at all. No, it is a big deal! But the big deal is missing from both $S = S_0$ or the separate $S_1 \cup S_2 \cup \dots \cup S_n = S$ and $S_1 \cup S_2 \cup \dots \cup S_n = S_0$ equations. The crucial thing is that in these separate equations, the relocation of the points is rigid! So picking three points from S_i , will determine the removal of the rest in S_i . Indeed, that's what rigidity means:



To compare this to a projection, let's regard the simplest of that, say a double stretch of $[0, 1]$.



Every point $.d_1 d_2 \dots$ becomes its double. It's quite easy to calculate this. Double them and write them upside down, then add the top 1 values to the previous:

$$.347690235 \dots \rightarrow \begin{array}{ccccccc} & & 1 & 1 & 1 & & 1 \\ . & 6 & 8 & 4 & 2 & 8 & 0 & 4 & 6 & 0 & \dots \end{array} \rightarrow .695380470 \dots$$

Or in reverse, from $[0, 2]$ we can divide any digit with two and write the remainders under the odd ones:

$$\begin{array}{ccccccc} . & 6 & 9 & 5 & 3 & 8 & 0 & 4 & 7 & 0 & \dots \\ & 1 & 1 & 1 & & & & & & 1 & \dots \end{array} \rightarrow .347690235 \dots$$

It is a very simple process, but all digits are increased or decreased, though not literally because of the remainders. A rigid shifting would merely mean adding a fix infinite decimal to all others. This of course would have to be done in a coordinate system for all three coordinates. And differently for the finite many pieces. So as we see, looking at the points as decimals, the reassembling is not even simpler than a projection. It's only the philosophical meaning is that makes the $S = S_0$ a stronger paradox than the projection.

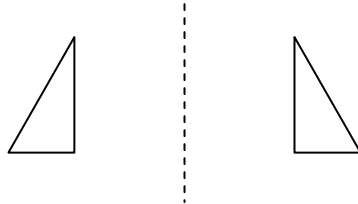
Our paradox at the end will be a sphere, splitting into three S, S_m, R disjoint subsets, so that S_m is an m mirroring of S to a diagonal of the sphere and R is merely a sequence. The two symmetrical main parts S and S_m will have rigid half collapses. This and the fact that the R sequence can be easily gained, that is duplicated, leads to a more dramatic consequence: The whole sphere can be duplicated from finite many subsets. Finally, using this layer by layer radially and just "create" an extra center, we can duplicate a full ball. So then the "magic" of creating volume, increases the effect even more.

62. Isometries of space

Before we go to the details of how to get rigid collapse, I have to tell the official name of rigidity, which is “isometry”. Iso = same, metry = size. So it makes sense.

More importantly though, we have to correct an over simplification in the name “rigid”.

To keep all internal distances is possible beyond the rigid displacement of a body, that is using shifts and turns. Namely, we can create a perfect mirror image too. That means mirroring all points to a fix plane. Such mirrored bodies cannot be displaced into each other, the same way as in two dimension, we cannot overlap by slidings, even just two mirrored triangles:



The only way we can cover them, is by removing one of them from the plane, then turn it around in space, and put it back into the plane.

So isometries are shifts, turns and mirrorings. The crucial trick to prove this, would be showing that every isometry has a line that remains in itself. Indeed, then it wouldn't be too hard to see that if this line is fixed, then we have a turn, if it is shifted, we have a shift, and if it is mirrored in itself, then we have a mirroring. Unfortunately, the conservation of a line is just as hard to prove. So the details are a pretty long exposition that I present here and you can just skip it over. Besides the fundamental conservation of a line, another interesting result is proved too at the end, namely that all isometries can be replaced by merely mirrorings:

D

PQ denotes the interval between P and Q.

$PQ = RT$ denotes that the two intervals have the same length.

$PQ \parallel RT$ denotes that the two intervals are parallel.

$PQ \nabla RT$ denotes that the two intervals are not parallel.

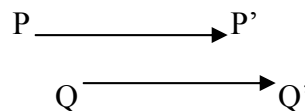
D

- 1.) A transformation is any P' assigned points to all P .
- 2.) $S' = \{P' ; P \in S\}$.
- 3.) P is fixpoint is $P' = P$.
- 4.) S is a fix set if $S' = S$.
- 5.) S is conserved set if $S' \subseteq S$.
- 6.) $P^* =$ middle point of PP' .
- 7.) $P_{\perp} =$ middle perpendicular plane of $PP' =$ perpendicular to PP' through P^* .
- 8.) The identity is the $P' = P$ transformation.

D

A P' non identity is:

- 1.) Isometry if $P'Q' = PQ$ for all P, Q .
- 2.) Shift if all PP', QQ' are:
 - a.) parallel
 - b.) same length
 - c.) same directional
- 3.) Mirroring to a Π plane, if for all $P, \Pi = P_{\perp}$.
- 4.) Turn with an α angle around an L line with one of its directions chosen, if P_L denotes the perpendicular projection of P to L , then $P'P_L$ is the α turn of PP_L in the plane perpendicular to L , looking from the chosen direction.
Of course, looking from the other direction, the angle would be $-\alpha = 360 - \alpha$.



T

Any P' transformation must obey exactly one of the followings:

- 1.) All PP' are parallel and all $P'Q'$ are parallel to PQ .
- 2.) All PP' are parallel, but not all $P'Q'$ are parallel to PQ .
- 3.) Not all PP' are parallel, but all P_{\perp} go through a fix L line.
- 4.) Not all PP' are parallel, and not all P_{\perp} go through a fix L line.

P

The four exclude each other and one must be true.

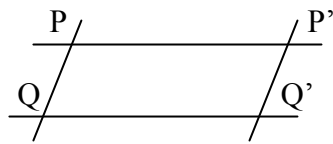
T

The four cases of the previous theorem for an isometry are:

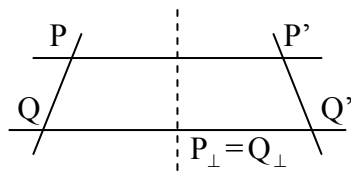
- 1.) Shift
- 2.) Mirroring
- 3.) Turn
- 4.) Turn then mirroring or turn then turn.

P

- 1.) All PP' are parallel by definition and $P'Q' = PQ \rightarrow PP' = QQ'$.



- 2.) Let $P'Q'$ be one that is not parallel with PQ . Then, $P'Q' = PQ \rightarrow P_{\perp} = Q_{\perp}$.

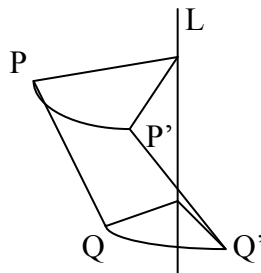


For any third R point, $R'P' \vee RP$ or $R'Q' \vee RQ$ must be true, so $R_{\perp} = P_{\perp}$ or Q_{\perp}

- 3.) If $PP' \vee QQ'$, then $P_{\perp} \vee Q_{\perp}$, so they cross in an L line.

We claim that P and P' and Q and Q' are not only equal distanced from this L line, but are turned with the same angle. That is, if P_L and Q_L denote their drop to L , then

$$PP_L P' \angle = QQ_L Q' \angle :$$



If all R_{\perp} go through L , then of course, this same turn works for all R to get R' .

- 4.) Let $PP' \vee QQ'$ and turn S around the L line, that is the crossing of P_{\perp} and Q_{\perp} with the α angle determined by P' and Q' . If this turned set is S_{α} , then it is isometric to S' and this $S_{\alpha} \rightarrow S'$ isometry has at least two fix points, namely P and Q .

If all $R_{\alpha} R'$ are parallel, then by 1.) and 2.), we have a shift or mirroring, but since we have fix point, we can't have a shift, so we have a mirroring. If not all $R_{\alpha} R'$ are parallel, then by 3.), we have a turn around the PQ line, because these are fix, so all R_{\perp} go through them.

D

Special isometries are:

The three basic: shift, mirroring, turn, and the following three basic combinations:

- 1.) A turned mirroring is a turn around an L and a mirroring to a plane perpendicular to L .
- 2.) A shifted mirroring is a shift and a mirroring to a plane parallel to the shift.
- 3.) A screw around L is a turn around L and a shift parallel to L .

T

In the combinations above, using “and” instead of “then” was justified because their order was immaterial. For example, the screw is a turn then shift or shift then turn.

T

All special isometries have conserved line.

P

Trivial one by one.

T

If an L line is conserved, then the isometry is special, namely:

- 1.) L is either fix or mirrored to a Π plane or shifted in itself.
- 2.) If L is fix then, P' is either a mirroring to a Π , containing L or a turn around L .
- 3.) If L is mirrored to Π , then P' is either the same mirroring in the whole space or a turned mirroring to Π around L .
- 4.) If L is shifted, then P' is either the same shift in the whole space or the same shifted mirroring to a Π containing L , or a screw with the same shift.

P

- 1.) If L has two fix points, then the whole L is fix.

If L has only one fix point O , then any other P must be mirrored to O to have $PO = P'O' = P'O$.

If L has no fix points, then for any P, Q on L we have $PP' = QQ'$.

- 2.) If all RR' are parallel, then since we have fix points, we can't have a shift, so it must be a mirroring to a Π and it must contain all fix points, including L .
If not all RR' are parallel, then we must have a turn around L because all R_{\perp} must contain all fix points, including L .
- 3.) Π is conserved too and either it is fix or for all R point of it, R_{\perp} contains L , so Π is turned in itself. If Π is fix, then in the space we have the same mirroring as in L , or if Π is turned, we have a turned mirroring.
- 4.) If R_{\perp} is the drop of R to L , then the RR_{\perp} distances are preserved. In other words, looking perpendicularly to L , we have an Θ plane in which P' is still an isometry. But here, L is a fix point, so we have the following possibilities:
 - a.) The whole Θ plane is fix, and thus, P' is the same shift in the space as in L .
 - b.) There is only a line fix in Θ , through L , which in space is a Π plane. Then, in Θ , we have a mirroring to Π and thus, a shifted mirroring to it in space.
 - c.) Only L is fix in Θ and then we have a turn in Θ and thus, a screw in space.

T

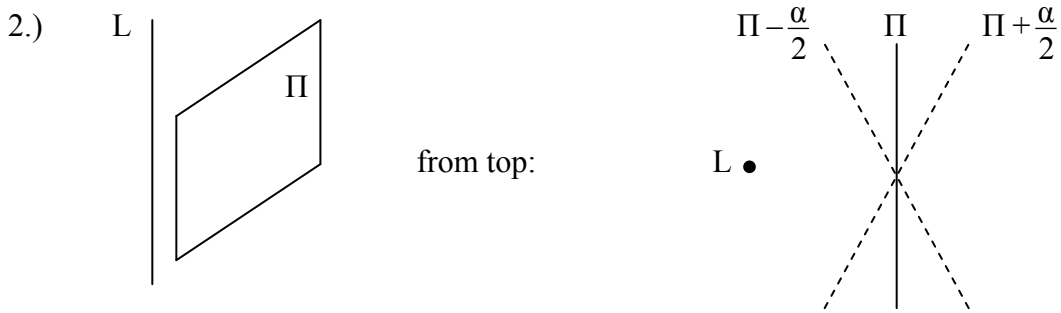
“Turn then mirroring” replaced by special isometries:

- 1.) A turn then mirroring to a plane going through the turn line is a single mirroring.
- 2.) A turn then mirroring to a plane parallel to the turn line is a shifted mirroring.
- 3.) A turn then mirroring to a plane crossing the turn line is a turned mirroring.

P



Thus, all RR' will be parallel and L is fix, so we have a mirroring.

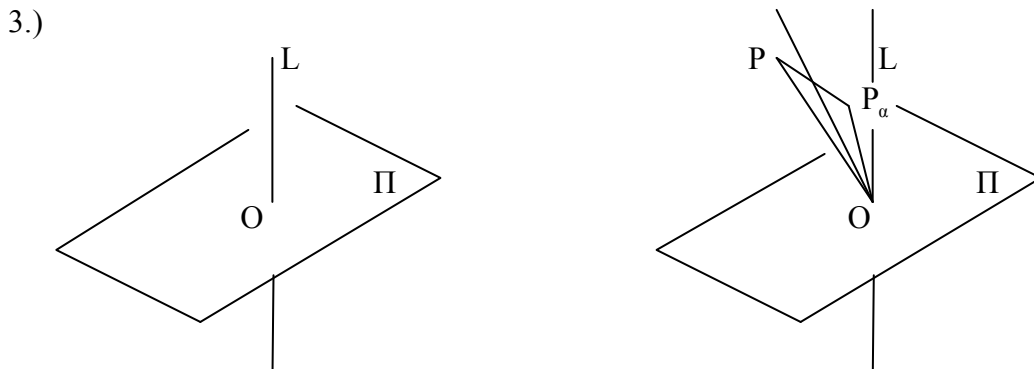


We turned Π with $\pm \frac{\alpha}{2}$. These are mirror of each other to Π and also, the α turn of $\Pi - \frac{\alpha}{2}$ is $\Pi + \frac{\alpha}{2}$. Thus, $\Pi - \frac{\alpha}{2}$ is conserved by the turn then mirroring.

If a line is conserved, it either has a fix point and is a mirroring or hasn't and is a shift.

Here $\Pi - \frac{\alpha}{2}$ can't have fix points, so it's shifted in itself, and thus the whole plane is a shifted mirroring to this same line.

Then, layer by layer, the whole space is a shifted mirroring to $\Pi - \frac{\alpha}{2}$.



By 3.) of previous theorem, enough to show that there is an L_0 line, that is mirrored in itself. The P point of the above picture defines such L_0 , if P and P_α are symmetrical to the perpendicular to Π . Indeed, P_α mirrored to Π will fall on the continuation of the $PO = L_0$ line.

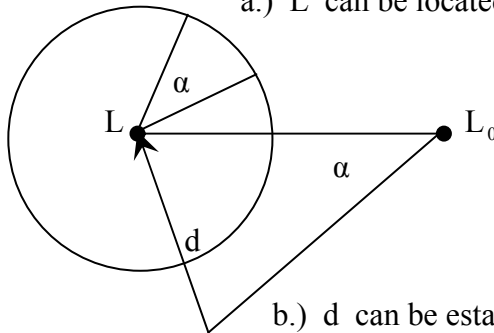
T

“Turn then turn” replaced by special isometries:

- 1.) a.) Every α turn around an L_0 line followed by a d shift perpendicular to L_0 , is just an α turn around an L parallel to L_0 .
- b.) For every α turn around an L line, and any L_0 line parallel to L , the turn can be replaced by same α turn around L_0 followed by a shift perpendicular to L_0 .
- 2.) A turn followed by a turn around an axis that is parallel to or crosses the first, can be replaced by a single turn.
- 3.) A turn followed by a turn around an axis, that is not in the same plane as the first, can be replaced by a screw.

P

- 1.) a.) L can be located as the point in the figure and it is fix:



- b.) d can be established from the figure and L becomes fix.

- 2.) There are two lines so that the first turn turns the first line in the second, while the second turn turns the second line into the first. Thus, the first line will be a fix line.
- 3.) Let L_1 be the axis of the first and L_2 of the second! Let L_0 be the parallel with L_2 that crosses L_1 . Then by 1.) b.), the turn around L_2 can be replaced one around L_0 followed by a shift perpendicular to them. Then by 2.), the turns around L and L_0 can be replaced by one around an L . The followed shift can be decomposed into a perpendicular and parallel component to L . The perpendicular shift melts into giving a new turn by 1.) a.).

T

- 1.) Every isometry is a special one.
- 2.) Every isometry has a conserved line.

P

Trivial by the previous theorems.

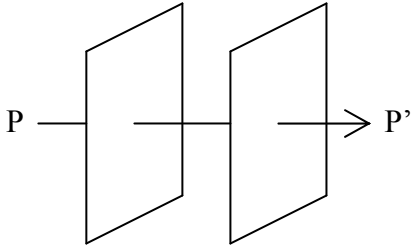
T

Every isometry can be replaced by a sequence of mirrorings, namely:

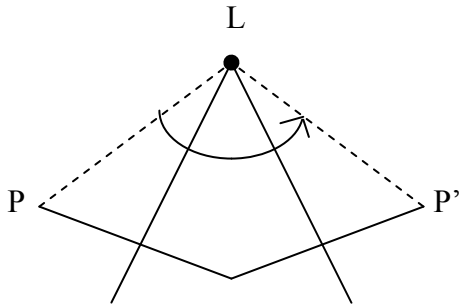
- 1.) A shift can be replaced by two mirrorings, both with perpendicular plane to the shift.
- 2.) A turn around L can be replaced by two mirrorings with planes crossing in L .
- 3.) Every special isometry is:
 - a.) one mirroring or
 - b.) two mirrorings or
 - c.) three mirrorings, where the last is perpendicular to the first two or
 - d.) four mirrorings, where the second two are perpendicular to the first two.

P

1.)



2.)



- 3.)
 - a.) A mirroring itself is basic.
 - b.) A shift is two parallel mirrorings by 1.).
A turn is two mirrorings with crossing planes by 2.)
 - c.) A shifted mirroring is two parallel mirrorings followed by a perpendicular one.
A turned mirroring is two crossing mirrorings followed by a perpendicular one.
 - d.) A screw is a turn then shift or shift then turn and thus is:
two crossing mirrorings followed by two perpendicular ones that are parallel or
two parallel mirrorings followed by two perpendicular ones that are crossing.

63. The mirroring paradox

The fact that a two dimensional mirroring can be replaced by taking one of the objects out into space, then turn it around and put it back into the plane, suggests that a three dimensional object's mirrored version could also be obtained by taking the object out into four dimension, turn it around and bring it back to our world. So a man's heart could go to the right side without operation or any intrusion.

The assumption of four dimensional creatures of course has seemingly even more amazing features too, like being able to take money out of a safe, without opening and so on.

But this mirroring without an interference with the body, is much more important.

The idea that we don't have to take apart the points and put them back oppositely, merely take a walk in a wider universe, is not that absurd after a while. In that wider universe, to be left handed or right handed is merely a choice of free will. Everybody can do everything backwards. This is merely a special case of the even greater abilities.

But the most amazing about all these thoughts is that there is a residue concerning our own world, regardless of the reality or impossibility of such physical walks to four dimension.

The only similar or maybe even bigger recognition is the relativity of time.

The interaction of things happening nearby created an illusion of consciousness, that time is universal. Relativity drilled a hole on this illusion, but didn't pursue a transformation of consciousness. Indeed, there are hundreds of "serious" articles on the internet, trying to explain the twin paradox, through the alteration of aging, due to a travel. Just like at the four dimensional creatures, here too, there are much wider consequences about what is fact, what is description, but the particular fact that by going away and coming back, somebody can become younger, than staying, is the unavoidable local end result. How we wrap our mind around it, is vital. Einstein didn't give a crap about what people think and how people think. In fact, he found it amusing that most couldn't accept the full meaning of relativity. It is hard to grasp that what we call aging, when it is measured by a single time interval, then it becomes relative. Most importantly, that this change in measurement has nothing to do with the internal process of aging, spending the time. So in this phrase, "spending the time," the time is not a fixed entity pre-existing of the spender. So the twin paradox should be broken down to many steps:

At first level, it is the paradox, that with universal clocking in different systems, that travel with fixed speeds, these universal times of the systems have to be different for each system.

The second level, is that all systems will regard the others as slower.

Third level, is that seemingly a local difference in opinions could not be created, because all systems have to keep their speeds.

Fourth level is that a return to a place can still be approximated, by going away, turning around and coming back.

Fifth level is that then, the two systems, the staying, and the leaving and returning, will again claim opposite youngening of each other.

Sixth level is that we can easily choose one of them as valid, namely the staying one, simply because it didn't break the restriction of fixed speed.

Seventh level, is that this judgement is actually correct, the staying twin will be right, so the other will become younger.

Eighth level is that the leaving and returning twin, must break the fixed speed restriction for only very short periods, namely at leaving to speed up, at turning and at slowing down in the end. So, the youngening should be caused by only these periods.

Ninth level is that on the contrary, the youngening formulas use only the long fixed speed trips, away and back. So, the mistaken twin has true formulas too, for his travel in the fix speed trips, that still turns out to be opposite, due to the minor other periods.

Tenth level is that then this whole view of youngening is meaningless! There is no youngening as a process, the trip is not causing youngening. The trip merely gives a method of calculating the difference of time.

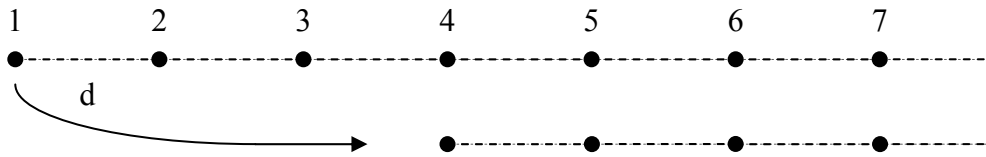
Now we can appreciate the mirroring paradox much better.

Having our heart on the other side after coming back from our walk to the fourth dimension, is the least of our problems. We'll look at the clock and get dizzy! Why is it going backwards?

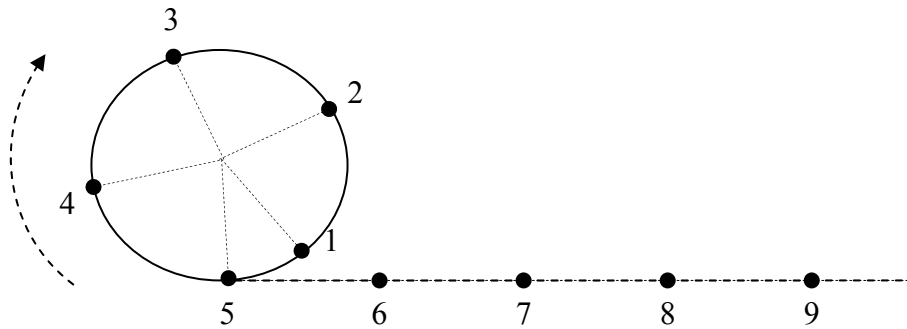
The numbers are correct, but start from the right. They go “anti-clockwise” in our mind. We pick up our guitar and we can’t play it. Did it turn left handed or what? No, we turned left handed. The fact that mirroring our brain causes instant mirrorings of memories, is a truth regardless whether we can mirror a brain or not. Math is a bigger reality than physics. People, wake up!

64. Bounded rigid collapse

Turning a bounded set into itself, can only be done if the set is already around an axis. But turning around a continuous object, like a cylinder or ball, achieves nothing. It just moves it around without collapse. Of course we can cut an object into subsets. So the weakest result towards our bigger goals is first, to find any non continuous set that can collapse at all. The solution is coming from the simplest unbounded case. The way the naturals collapse into themselves.. But now, the equal distance is emphasized too.



Now lets roll the full set onto the c circumference of a circle.

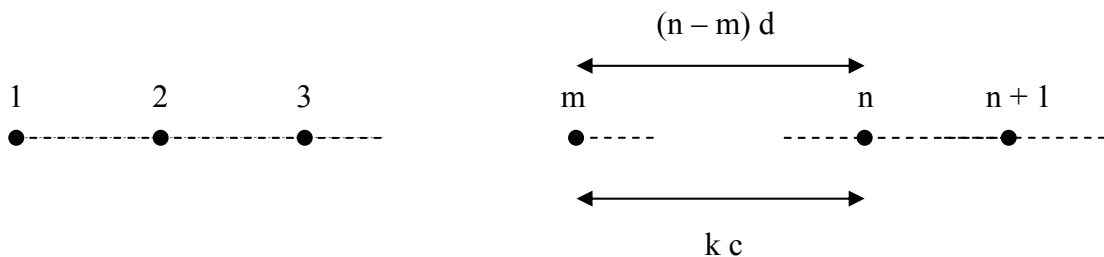


All we have to make sure is that no new rolled up n number is falling onto an earlier m .

By the way the d distance on the circle corresponds to $\frac{d}{c} 360^\circ$ angle.

But can we avoid collision of the rolled up points?

Lets see what a collision of n with m on the circle would mean back on the line:



But: $(n - m) d = k c \iff \frac{d}{c} = \frac{k}{n - m}$

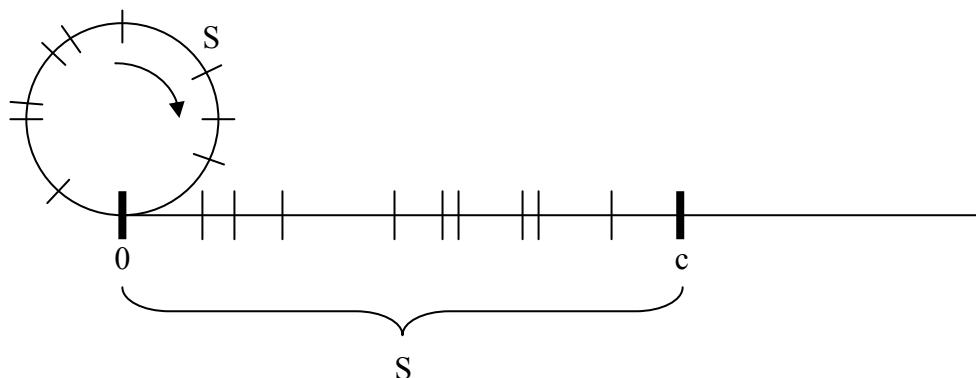
Thus, if the d distance of the naturals and the c circumference of the circle are chosen so, that $\frac{d}{c}$ is irrational, then such collision cannot happen.

Then a $\frac{d}{c} 360^\circ$ turn will move the set on the circle exactly as a d shift does on the line.

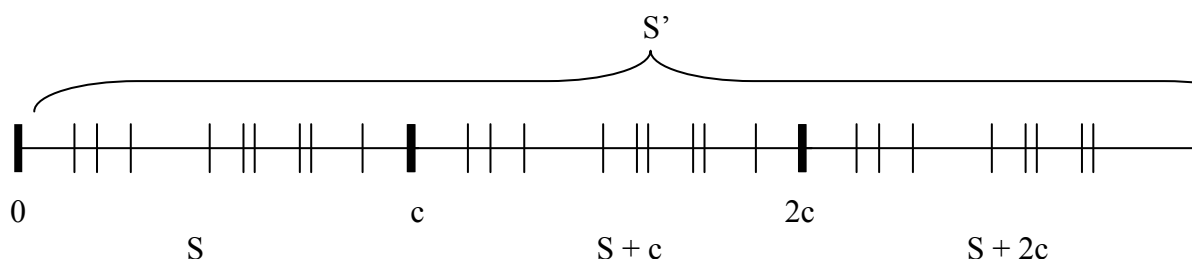
So we lose 1. Using an m multiple, that is $m \frac{d}{c} 360^\circ$ angle, we lose $1, 2, \dots, m$.

65. "Losing an S set from a circle"

I used quotation marks, because we can't do this, we merely aim for it.
The idea is simple. We pick an S set on the circle, and roll it onto the line.
So from the circle, we get an imprint of the same S set in $[0, c]$



But we keep on rolling though, and thus, we'll get a c periodic, unbounded S' set:



Can we shift S' away with a d distance, that is form $S' + d$ so that it wouldn't bump into itself? This seems pretty hard, but we aim even higher. We'll shift S' also repeatedly, that is with $2d, 3d, \dots$ too and require that all these $S', S' + d, S' + 2d, \dots$ are disjoint.

The total $T = S' \cup S' + d \cup S' + 2d \cup \dots$ can be looked in another way too.

Namely, first forming $S^* = S \cup S + d \cup S + 2d \cup \dots$ as a disjoint shifts of S by d , and then the c shifts of this, that is $T = S^* \cup S^* + c \cup S^* + 2c \cup \dots$.

Indeed, the bottom line is that for no two $P, Q \in S$ can $P + j c + k d = Q + m c + n d$.

But the c periodicity of S^* means that rolling back T to the circle, it is merely S^* .

Thus, on the circle, S^* moves in itself when turned with $\frac{d}{c} 360^\circ$ and so S is lost.

So we couldn't guarantee a turn of the full circle, with losing S , only the turn of an S^* subset of the circle containing S . All this of course, must depend on S , which is obvious because if for example S is the full circle, then there can't be such S^* .

So what is the necessary condition for S to have an S^* ? Very simple, S has to be a sequence!

Now we might think that then the proof of this will rely on S being a bounded sequence on the line, and S' being a periodic repeatance of it.

Obviously S^* is a sequence too, because a sequence of sequences always is.

The beauty of our proof will be that the bounded S and its periodic repeatance is immaterial.

For any S' sequence on the line, there is a d that $S', S' + d, S' + 2d, \dots$ are all disjoint.

In fact, we claim that a d distance exists too, that all $S' \pm \frac{m}{n} d$ are disjoint.

This is amazing, because it means that not only S' can be a dense set on the line, but that the

$\pm \frac{m}{n} d$ shifts are also dense, namely the rational multiples of the d unit.

So a dense set can be densely shifted without coinciding at all.

This shows how much room is in the full continuous line. Most amazingly, the proof also shows that it has nothing to do with density. It's all about sequencabilities. Indeed, first of all:

The $\pm \frac{m}{n}$ rational moves are merely a sequence of d_1, d_2, \dots shifts. So:

T For any P_1, P_2, \dots sequence of points and d_1, d_2, \dots sequence of shifts, there is a d distance, that using the dd_1, dd_2, \dots shifts, no P_m goes into P_n .

That is: $P_m + dd_k \neq P_n$

P This of course is the same as not having any P_m, P_n points that have dd_k as difference. The possible $P_n - P_m$ differences are themselves just a p_1, p_2, \dots sequence of shifts. So what we really claim is that for any two p_1, p_2, \dots and d_1, d_2, \dots sequences of shifts, there is a d that $p_m \neq dd_n$ that is $\frac{p_m}{d_n} \neq d$.

But indeed, the possible $\frac{p_m}{d_n}$ ratios are merely an r_1, r_2, \dots sequence again.

So all we claim is that for any sequence of real numbers, there is one not among them.

This of course, is obvious by the non sequencability of the real numbers.

Thus, the incredible depth of the line, where dense subsets can move densely without collision, is merely a direct consequence of its non sequencability.

66. Half collapse. My favourite paradox

The S set was any sequence on a circle and we created an S^* wider sequence, that can be collapsed, losing S . That is, S^* going into $S^* - S$. Then of course, $S^* - S$ is a sequence too, and due to our little detours above about densities, we might hastily think that we not only achieved bounded collapse, but bounded half collapse. But this is far from true. It's true that S and $S^* - S$ are both equivalent sets, but not the same rigidly.

In the final promised break up of the S sphere into $T \cup T_m \cup R$ both T and T_m will be half collapsing sets. Amazingly, quite directly, T will turn into two halves of T_m and T_m into two halves of T . The only simpler case could be if a set would directly turn or shifted into its two halves. Among bounded sets, this cannot happen, but in an infinite plane it can.

This is what I call real magic. Not just because it's not a language increased or verbally hyped up form, like the doubling of spheres or balls, but also because the example that produces it is totally direct. So it is a seemingly absurd visual possibility, yet directly visualized.

A real widening of our intuitions.

The drawback is that it can only exist for unbounded S set, and so it is seemingly not part of the promised ball doubling circus. That's why it is not even mentioned usually.

Finally, it was discovered by one of the few mathematicians that I respect as early representatives of the anti Formalist tendencies, that by today has almost died out.

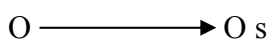
A real nice guy called Sierpinski. Quite an opposite of the more famous Tarski, who discovered the doubling paradox, and was a producer of more Formalists than any other "educator" in the estranged world of American university mathematics.

So here it is, the Sierpinski-Mazurkiewicz paradox.

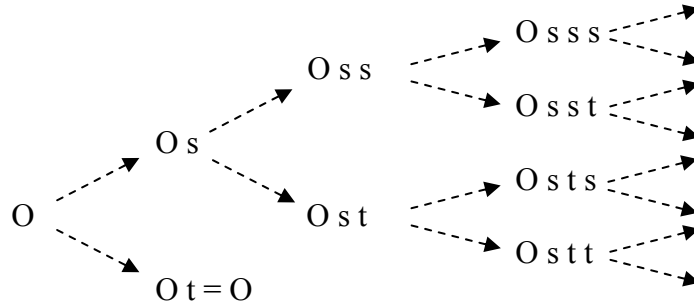
S will be a point set in the plane. Since it directly collapses into S_0 and $S - S_0$, these motions will be the two simplest moves of the plane, an s shift and a t turn:



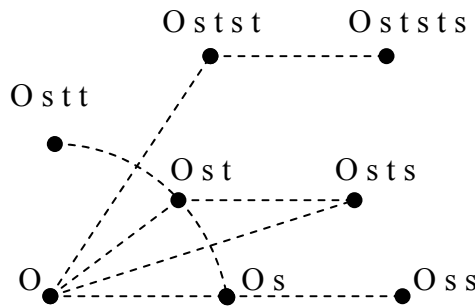
Since t had the fix O around which it is turning, we should start with O as element of S . Applying t to O is useless, $O t = O$. Quite on the contrary, $O s$ is a new point:



The $O s$ can again be shifted or turned to obtain $O s s$ or $O s t$. Then again, and again, we can create two-two points, so actually, we have a dual branching tree:



In fact, the actual points in the plane look very similar to these schematic branchings:



Of course, it becomes messier and messier, especially if you think about that the turns eventually “add up” to turns around to be even more than 360° .

The only real concern about all this is that a new point might bump into an old.

We also see that the chosen α angle of the t turn is the only crucial factor. The length of S is immaterial, and we also feel that even if an α angle is coinciding, that is leads to a point that is an earlier already, we can alter α to avoid this. To be more precise, we can regard any finite long: $O s t t s s s t t s \dots t = P$ hypothetical branching point, and visualize it as actual point in the plane with all possible α angles from 0° to 360° .

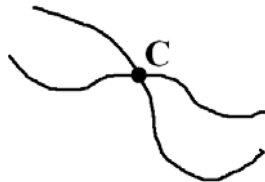
That is, P becomes a $P(\alpha)$ curve.

At $\alpha = 0$ all the t -s are useless, so $P(0)$ is merely using all the s shifts:

$P(0) = O s s \dots s$. The same is true at the end value, that is for $P(360)$.

In between, $P(\alpha)$ will go around in loops. Now, can it collide with an earlier, that is shorter branching point Q ? Well, it depends on what that was. But, it had to be one made by the same angle. So, we can make the same $Q(\alpha)$ curve for Q and check where $P(\alpha) = Q(\alpha)$.

These are simply the crossing points of the two curves. Visually, it is obvious that the two curves, $P(\alpha)$ and $Q(\alpha)$ can only cross finite many times. But this actually relies on our belief that once they cross, they go differently.



If that's the case, then the C crossing has a surrounding, where there is no new crossing.

In short C is isolated among the crossing points. Then obviously, there can only be finite many crossings in a bounded space due to Bolzano-Weirstrass and closedness of the curves.

The weird possibility that two $P(\alpha)$, $Q(\alpha)$ curves could cross densely is quite hard to refute.

So lets assume that we only have finite many coinciding angles of $P(\alpha)$ and $Q(\alpha)$.

Then of course, the total of all coinciding angles is merely a sequence, so obviously we can

choose good, never coinciding α . So the branchings in the $O s \dots$ sequences in our tree, perfectly correspond to different points in the plane. These points will be our claimed S set. And now comes the big trick. Let S_0 be all the points that were created by $O s \dots s$ sequences, that is ones ending with a shift. Then of course, $S - S_0$ is the $O s \dots t$ ones, that is the turn ending ones, plus the missing O itself, that has no ending at all.

Then, $S s = S_0$, $S t = S - S_0$.

What the hell is $S s$ and $S t$? Well, applying the s shift and t turn, to the whole S set.

We generated the S set as O plus all the possible finite $O s \dots$ sequences, realized in the plane. So then, $S s$ is exactly the same as adding an s to O , plus to every $O s \dots$ sequence. These are exactly the s ending generations. What is $S t$? It is the t turning of the full set, that is the t turning of O , which is itself, plus all the $O s \dots$ sequences, which becomes exactly, the t ending generations.

Our infinite S point set, will rigidly shift and turn into itself, namely into two disjoint subsets.

So the naturals, moving into the evens or the odds, is reproduced here in the plane, but rigidly.

Whereas the naturals on the line, would have to stretch when moved into the evens or odds.

67. The leap to continuity

The unboundedness and the two dimensionality of the Sierpinski paradox seems quite unavoidable, due to the basic shift, turn applied in it. But another feature we aim for, continuity, doesn't seem that alien to it. Anyway, a bad idea is always the best step towards the good ones.

So, lets pursue a generalization of our S set above in the plane. Formalists of course, never travel in bad directions, to keep the students totally paralyzed and void of asking questions themselves. So here is the simple idea, if O and s and t created S , then why don't we simply use a new $O' \notin S$ and create a new S' starting from O' . Then, pick again a new O'' to make S'' , and so on. At the "end", which is a big leap, there would be no more points left, so: $S \cup S' \cup S'' \cup \dots$ were the whole plane.

So then, the whole plane could be shifted or turned into its disjoint subsets.

But this is a total nonsense, both a shift and a turn of the whole plane is the whole plane itself.

So where did we go wrong? We might think that at the uncomfortable leap, that is by over simplifying the disappearance of all points. In fact, the non sequencability of the continuum, even tells us that a simple $S \cup S' \cup S'' \cup \dots$ union of a sequence of sets, can't be a continuum, because each was a sequence itself. Yet, that's not the problem. Or rather, that's not the real problem. This picking and continuing doesn't have to stop after S, S', \dots

In other words, we can continue a sequence with new sets.

It is a leap and it is a strange one, but it didn't cause the contradiction.

We simply made a false assumption already at S' , that it will lead to new points. This happened because we were so meticulous about showing that the points in S can be generated without collisions. Actually, there was still some unfinished business about the finiteness of the

$P(\alpha) = Q(\alpha)$ collisions. But all this is not the same as a possible collision of points started from O and O' . Clearly, similar arguments can work here too. So we can only have a sequence of collisions, and thus indeed, we can pick O' without collision. But this, then alters the free pickings of newer and newer O -s. So in a sense, it was still correct that the continuation leap was involved.

This raises the question, whether there is a situation of points and motions, so that the new pick definitely can be done freely. Amazingly, this is very easy. All we need is to include the reverse of the basic motions. For example, using $s, t, -s, -t$ and again generating all possible finite sequences from these. Then, if we pick a new O' start, then :

$$O \sqrt{s s t - s t - s - t} = O' \sqrt{-s t - s t s - t}$$

is impossible, if already the O sequences were different.

Indeed, applying the reverse of the sequence used after O' to both sides, we would get:

$$O \overbrace{sst-st-sts} \overbrace{t-s-ts-ts} = O' \overbrace{-st-sts-t} \overbrace{t-s-ts-ts} = O'$$

$\underbrace{\hspace{10em}}_{\emptyset}$
 $\underbrace{\hspace{5em}}_{\emptyset}$
 $\underbrace{\hspace{2em}}_{\emptyset}$

So, O' would turn out to be not a new point at all.

Before we jump to joy, that after all we can collapse the full plane into itself, and thus find a contradiction in mathematics, we should realize that the $-s, -t$ added moves, completely ruined the original idea of splitting the total set, by the endings in the generations.

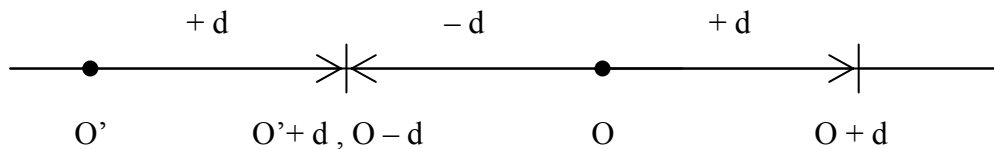
Indeed, applying for example, s to the $O s s t - s$ point, we actually get $O s s t - s s = O s s t$,

so a t ending one. We might counter argue to use more distinctive applications, and indeed, this will be the heart of the final doubling paradox. But now lets apply the reverse complete motions idea, for a new direction, without attempting sub collapse at all.

We go back to the infinite line where the rationals can amazingly be moved away, rigidly to different also dense locations, without having any coinciding points in the different moved versions. Then I explained that this whole density surprise is only subjective, because it was a mere consequence of the sequencabilities. But now we don't want to stop at using the rationals or any other sequence to be shifted to new locations. So now, again not the density will be the real importance, rather that the shiftings contain the reverses.

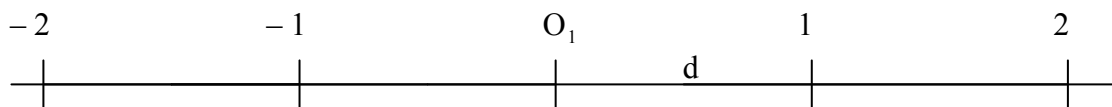
For rational multiples of the new unit d , that is for $\pm \frac{m}{n} d$, it was unimportantly true then.

Now, it is the only importance. This then includes the much simpler case of $\pm m d$ moves. Observe that this is the simplest case, because we have to have the repetitions included too. Indeed, I said we don't pursue the sub collapse, but the repeated move sequences are needed anyway, because our logic of reversal eliminations, used it too. For example, the simplest reverse complete move set is $\{d, -d\}$. But then, picking a new O' after the first set, $\{O, O + d, O - d\}$, we can't guarantee that not only O' is new, but its moves too. Indeed, if $O' = O - 2d$ then $O' + d = O - d$.

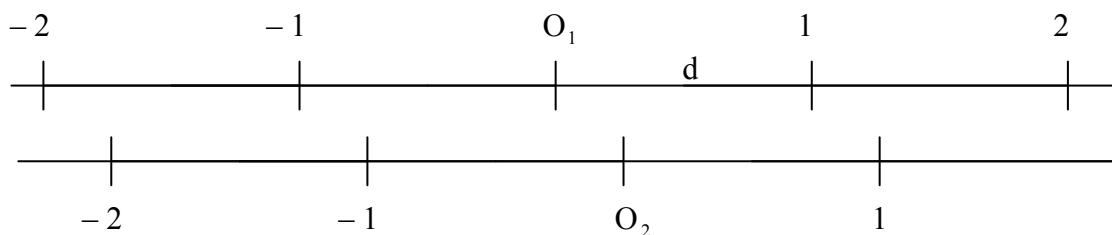


Now using the simplest move set $\{\pm d, \pm 2d, \pm 3d, \dots\}$ we can start the heuristic "pick to the end" process.

O_1 is any point of the line. The moved ones of O_1 are the \pm whole numbers with d as unit:



Then pick a new O_2 and form the wholes starting from that too:



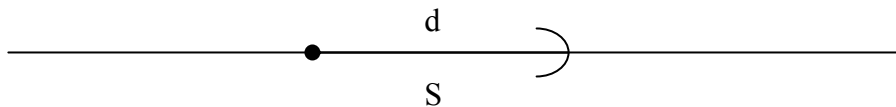
The big thing is that there are no strings attached. Any new point can be picked and then all of its \pm wholes will be new too.

After infinite many O_1, O_2, \dots we only have a sequence of sequences, thus a sequence in total, and so there will be definitely points on the line not picked or moved yet.

The picked new O after the sequence can be denoted as O_ω then $O_{\omega+1}, O_{\omega+2}, \dots$ and so on. The notation is unimportant, the main thing is that for any S set of already picked O -s, there is a next, unless S exhausted the full line. If that's the case, then the S set will be such that itself and its moved versions cover the whole line, that is:

$$S \cup (S + d) \cup (S - d) \cup (S + 2d) \cup (S - 2d) \cup \dots = L = \text{line} \quad \text{Can this be?}$$

Of course it can. The simplest example is any full d interval, but only with one end included:



Easy to check, that this and every single of its $\pm d$ shifted versions are disjoint, and in total they cover the line. To get this S as an end result of blindly picked points, is very unlikely of course. And that's the heuristic new feature! We have a method, without methods.

The particular truth that an intentional design can be the same as the random selection, is good. It shows that we are not crazy, that the picking idea is reliable.

Now we can then say, that using other move sets, that are repetition and reverse complete, we always must have a final S set, so that S and its moved versions are disjoint and cover the line. Even if for these other moved sets, we can't find easy methodical examples.

The simplest such case is if the moved set is the rationals, or at least the fractions with certain type of denominators. The 2 powers are good, because sums of such 2^k denominator fractions, are again the same: $\frac{m}{2^k} + \frac{n}{2^k} = \frac{m(2^{k-k})}{2^k} + \frac{n}{2^k} = \frac{M}{2^k}$

Since there are arbitrary small moves in our move set, here we definitely couldn't get an interval as S . In fact, S can't contain a full I . Indeed, S moved with any of the moves, must become disjoint, but any inner point of an I , moved with small enough shift, still stays in I .

Amazingly, in spite of this, S can become a set, contained in arbitrary small I interval. And most amazing is, how easy this is to see. All we have to do, is try! Namely, we restrict our blind pickings from I . Now the only problem could be, if we ran out of points in I before, on the whole line. But that's impossible! Indeed, if a P is on the line, not picked yet, then no moved versions of P was used either. Since, we have arbitrary small moves, with repeated use of such, we can move P into I . So we can pick this.

The possibility of an S in arbitrary small I , might confuse someone and say, "I remember this, it means that S is a nil set." This is way off though! First of all, a nil set is covered by pieces of an arbitrary small interval, not with the single interval. The only thing that can be covered by arbitrary small single interval, is a point. More importantly though, here not the S sets each are covered by arbitrary small interval, rather there are arbitrary small possible S outcomes.

So we pick first any I and then trying lots and lots of pickings, one must come out inside I .

This still leaves the possibility that the mistaken reaction was correct, that is the S outcomes are all nil sets. The truth is the opposite! They can never be!

Again, to see it, is shockingly simple. If an S were coverable by pieces from an arbitrary small I_0 , then first of all lets sequence all of our moves as m_1, m_2, \dots . This is easy, because the fractions can be sequenced. So then, the moved versions of S are a corresponding sequence: S_1, S_2, \dots . But $S \cup S_1 \cup S_2 \cup \dots = \text{full line}$. This is a contradiction.

Indeed, let I be any interval and cut it into $\frac{I}{2}, \frac{I}{4}, \dots$. Then, use each of these, as I_0 for

S, S_1, S_2, \dots . That is, cut each $\frac{I}{2}, \frac{I}{4}, \dots$ into pieces, to cover S, S_1, S_2, \dots .

Then the total pieces of I would cover the whole line.

68. Putting the genie on the silver platter

It's obvious that we let out a genie. The S set above with its three features:

1. never containing full I ,
2. possible as outcome inside any small I ,
3. never being coverable by arbitrary small I_0 ,

is very hard to conceive. So let's use the decimal system and see if S is "real" or not.

It will turn out to be very real.

Instead of the above 2^k denominatored fractions as moving set, let's use the 10^k ones, that is exactly the decimals. The numerators can be anything, so we can obtain any finite decimal with a whole part too. For example, $3\ 6\ 0\ 5\ 8.\ 0\ 2\ 3\ 5\ 1\ 2$ means $\frac{3\ 6\ 0\ 5\ 8\ 0\ 2\ 3\ 5\ 1\ 2}{1,000,000}$.

Shift to the right means adding, to the left means subtracting $3\ 6\ 0\ 5\ 8.\ 0\ 2\ 3\ 5\ 1\ 2$ from any point. The points themselves are all possible \pm decimals, but infinite after the decimal point.

So then shifting a point, simply means changing the beginning in an infinite decimal.

Beginning here means any short or long section and has nothing to do with the decimal point.

Now comes the reality shock!

Pick any infinite decimal as O_1 ! To form the shifted versions of O_1 , we have to add or subtract every possible finite decimal from O_1 . So in short, change its beginnings, in every possible way. The simplest is to cut off its first digit, change it in all ways. Then cut off two digits, change the second for every first choice, and so on. We change everything, but the "end" is always the same. In a precise way: From a point on, it always remains the same.

So picking the new O_2 is easy! We have to make an infinite decimal, that is different towards infinity too. In short, we have to change infinite many digits. Then, we apply again the beginning variations to this. Then pick a new infinite one, and so on.

Looking this way, it's quite easy to see what the S set of the O_1, O_2, O_3, \dots new picks will become. It's a collection of infinite decimals, so that every possible ending is there, but only one for each. In short, S is a representation from all possible decimal endings.

Changing any element of S , only in its beginning, is giving a perfectly legitimate alternative of S , still representing perfectly all endings. This at once explains the three features:

1. A full interval in S would mean to have all points between two points, that is between two infinite decimals. Now let's regard the first digit, where they are different.

$$\begin{array}{l} 3\ 6\ 0\ 5.\ 3\ 0\ \boxed{2}\ 3\ 0\ 5\ \dots \\ 3\ 6\ 0\ 5.\ 3\ 0\ \boxed{3}\ 2\ 1\ 9\ \dots \end{array}$$

I picked an example where the smaller is merely 1 less in this digit than the bigger.

But even here, we can safely change the next digit 3 to a bigger. Thus, we at once obtain same endings with different beginnings. They shouldn't be in S .

2. Any small I can be narrowed by finite decimals, and then staying between these beginnings, we can still get all endings.
3. The shifts are all decimal fractions. This is at once a double sequence namely the whole part shifts to the unit intervals and the decimal part inside each to the exact location. Together it is still just a single sequence and so covering the shifted versions with smaller and smaller lengths, we could cover the whole line with an interval.

69. Doubling a ball

Now that we showed how weird the step by step picked S sets can be, we have to return to the promised final result, the Banach Tarski paradox or the doubling of a ball.

Just to repeat again, this claims that there are finite many disjoint subsets together giving the whole ball, that is $B = S_1 \cup S_2 \cup \dots \cup S_n$, so that some of these sets can be reassembled into a full B and the rest of them into an other one.

The crucial method to achieve this is the Hausdorff paradox, decomposing a sphere into three parts as $S \cup Sm \cup R$ where m is a mirroring. R is merely a sequence and the S and Sm mirrored “halves” both have rigid half collapses. Namely a single t turn moves S into the St subset of Sm and double usage of t that is Stt will give the other half.

So $Sm = St \cup Stt$. Then of course similarly $S = Smt \cup Smtt$.

This indeed means that the four pieces: St , Stt , Smt , $Smtt$ can be reassembled into almost two spheres as $(St) \bar{t} \cup (Smt) \bar{t} = S \cup Sm$ and $(Stt) \bar{t} \bar{t} \cup (Smtt) \bar{t} \bar{t} = S \cup Sm$.

Here \bar{t} meant the reverse turn of t .

In fact, t will be 120° that is one third of a full turn, so actually $\bar{t} = tt$ and $\bar{t} \bar{t} = t$.

Of course, the two missing R -s in these doublings must be created from the original single R .

This can be easily achieved, because we showed how to eliminate a sequence on a circle.

Finally, the layer by layer application of this doubling combines into “radial” sets in a ball.

The only missing part is to double the center of the ball, which can be easily “stolen” from these R sequences.

So the basic idea is to use the m mirroring or 180° turn and a t 120° turn in place of the shifts and turns that we used in the plane. The big difference now, is not that both moves are turns, rather that we aim for not only repetition complete but reverse complete move sequences. We need this to ensure the continuability. And indeed the reverse of m is itself, while of t is tt .

This means that the possible sequences are much simpler. Basically the m -s and t -s must alternate but t can repeat twice. Also we can start with m or t , so all move sequences are merely beginnings of the infinite sequence: $(m) t (t) m t (t) m t (t) \dots$

The bracketed members are usable or can be skipped over.

The obvious three classes of the move sequences are the m , t or tt ending ones and we abbreviate them as $\{\dots m\}$, $\{\dots t\}$ $\{\dots tt\}$.

Applying a single move to $\{\dots m\}$ means applying the move to all elements and then:

$\{\dots m\} t = \{\dots t\} - \{t\}$. Indeed, we get all t ending ones, except the single t move.

$\{\dots m\} tt = \{\dots tt\} - \{tt\}$ similarly.

$\{\dots m\} m = \{\dots t\} \cup \{\dots tt\} \cup \{I\}$ here I means the identity, that is not moving.

Indeed, applying m eliminates the previous, but for the (m) start choice, gives nothing.

Thus, aside from the missing t and tt at the turnings, and the extra I at the mirroring, we got a very paradoxical $\{\dots m\}$ set of moves, because its mirroring is the same as the two turnings of it together. Amazingly, to get rid of the minute errors of t , tt , I is quite involved:

A natural idea is to add I to $\{\dots m\}$, because then t and tt are at once obtained by the turnings of $\{\dots m\} \cup \{I\} = \{\dots m, I\}$. But this brings even more trouble, because then the mirroring of $\{\dots m, I\}$ will contain m . So then this should be included in $\{\dots t\}$ or $\{\dots tt\}$.

This shows that a workable correction must involve infinite many changes in our basic three sets, so the simple m , t , tt ending classifications become a messy affair.

And yet, there is a way to make this change:

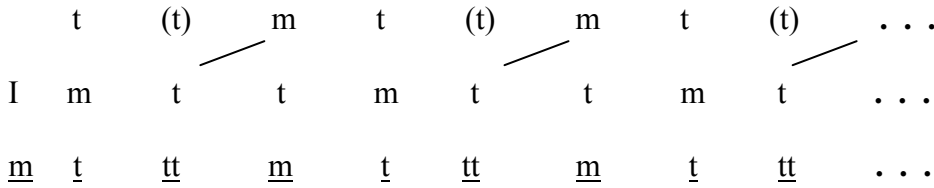
We first define the t starting move sequences exactly like above, that is we simply omit the possible (m) start and only regard the $t (t) m t (t) m t (t) \dots$ infinite sequence.

The beginnings are in three classes, now denoted as \underline{t} , \underline{tt} , \underline{m} according to the last elements:

t	(t)	m	t	(t)	m	t	(t)	\dots
\underline{t}	\underline{tt}	\underline{m}	\underline{t}	\underline{tt}	\underline{m}	\underline{t}	\underline{tt}	\dots

The lower members denote the sequences above them upto that point.

For example, the first $\underline{m} = t m$ or $t t m$, the second $\underline{t} = t m t$ or $t t m t$ and so on. Now we extend these classes with the help of an additional second sequence, that contains I and the m starting parts of the finite sequences upto the first single t usage. This can be easily achieved by having branching possibilities into the t starting sequences:



Again, \underline{m} , \underline{t} , \underline{tt} denote the sequences upto that point. So quite explicitly:

$$\underline{m} = I, t \dots m, mtt, mtt \text{ repeated}, m \dots t \dots m, \text{ (a single } t \text{ must appear)}$$

$$\underline{t} = t, t \dots mt, m, mtt \text{ or repeated} + m, m \dots t \dots mt$$

$$\underline{tt} = tt, t \dots mtt, mt, mtt \text{ or repeated} + mt, m \dots t \dots mtt$$

Now we truly have: $\underline{m} t = \underline{t}$ and $\underline{m} tt = \underline{tt}$ and $\underline{m} m = \underline{t} \cup \underline{tt}$.

Indeed, observe that each set has five types of members. The t and tt applications to \underline{m} are exactly underneath in the five types. For the m application to \underline{m} observe:

$$I m = m \in \underline{t}$$

$$t \dots m m = \begin{cases} t m m & = t \in \underline{t} \\ t t m m & = t t \in \underline{tt} \\ t \dots m t m m & = t \dots m t \in \underline{t} \\ t \dots m t t m m & = t \dots m t t \in \underline{tt} \end{cases}$$

$$m t t m, m t t \text{ repeated} + m \in \underline{t}$$

$$m \dots t \dots m m = \begin{cases} m t m m & = m t \in \underline{tt} \\ m \dots t \dots m t m m & = m \dots t \dots m t \in \underline{t} \\ m \dots t \dots m t t m m & = m \dots t \dots m t t \in \underline{tt} \\ m t t \text{ repeated} + m t m m & = m t t \text{ repeated} + m t \in \underline{tt} \end{cases}$$

Exactly the five five types of \underline{t} and \underline{tt} were obtained.

Applying these three sets of moves to a P starting point, we get three point sets,

$$P \underline{m}, P \underline{t}, P \underline{tt}. \text{ If } P \underline{m} \text{ is mirrored, it goes into the other two: } (P \underline{m}) m = P \underline{t} \cup P \underline{tt}.$$

$$\text{If } P \underline{m} \text{ is turned by } t \text{ or } tt \text{ it goes into them separately: } (P \underline{m}) t = P \underline{t}, (P \underline{m}) tt = P \underline{tt}.$$

Thus of course, $P \underline{t} \cup P \underline{tt}$ is a set with half copy. Namely a mirroring and one turn gives the $P \underline{t}$ half, while a mirroring and two turns gives the other $P \underline{tt}$.

Most importantly, new P can be chosen, again and again. Thus, we obtain an S set so that S plus its mirroring would be the whole sphere. One problem is that the turns should never coincide with mirrorings. This will stand if the two diagonals around which m and t are applied, have proper angles. For example, 45° is suitable. The other problem is that even so, the end points of the diagonals are not moving, so their mirrorings and turns must be excluded from the sphere. These are the R remaining sequence from $S \cup S m$.