Numbers

Steven Pemberton

Amsterdam

For *u*i*n

1 2 3 4 5 6 7 8 9

Numbers

Numbers are an idea, not a thing: an *abstraction* we would call them.

That is because there is no *three* that you can point to, only three *somethings*, such as three sheep, three trees, three balls. The idea of *three* is what those sheep, trees and balls have in common.

Because *three* is an abstraction, we need to have some way to make that idea real. We do this by *representing* a number in some way.

One way is by writing the name: *three*.

But we have other ways, for instance 3, as we write it, or III as the Romans wrote it. Or 三 as the Chinese write it. These are all representing the same number.

Note that all three of these representations look (a bit) like three stripes.

That's probably because originally people started writing numbers down as stripes.

Addition

Addition is easy. You want to add numbers: you write a number of stripes for the first number, and then a number of stripes for the second, and count up how many stripes you have: // plus /// is /////. This is how children learn addition, except using fingers for stripes. This is a *procedural* definition: it tells you how to calculate the result.

We write addition as `a+b`.

There are a number of *laws* to do with addition. The most obvious one is

```
a+b = b+a
```

This is called the *commutative law*, and it says that with respect to addition it doesn't matter which way round the operands are, 2+3 will give you the same result as 3+2.

Clearly //+/// is the same as ///+//.

This is a different use of the word *law* than is usual understood in everyday life. Usually laws are things created by people to help run society. In maths in means something that is always true. It might be better to call it a *fact*.

Subtraction

The complementary operation to addition is subtraction, which we can define as

```
(a+b)−b = a.
```

This may look like an odd way to define something, but it is called a *declarative* definition: it doesn't tell you how to calculate the result (although we do learn how to do that at school), it just tells you how to recognise the right answer.

In fact we use declarative definitions a lot. Probably the first time we are taught such a definition is for square roots. We are told that

> *the square root of a number n, is a number r such that* $r \times r = n$.

This doesn't tell us how to calculate it (and we have calculators to do that for us), but it does allow us to understand what a square root is, and how to recognise one.

What does this definition for subtraction mean?

If you have a number that is the result of adding `a` and `b`, and then subtract `b` from it, we get `a` back.

So, if we want to calculate

```
9−3,
```

then `(a+b)` is 9, and `b` is 3, so we have to find the value `a` such that

```
(a+3) = 9
```

which is 6.

So subtraction means “find the value such that adding the second operand to it gives us the first operand”.

You might prefer it if I say: imagine that the notation

```
[a+b]
```

means “a number that is the result of adding a and b”, for instance 9. Then

```
[a+b]−b = a
```

is true for all possible values of a and b.

So with 9−3

```
[9]−3 = a
[a+3]−3 = a
[6+3]−3 = 6
```

Since addition is *commutative*, we also have

```
(b+a)−a = b
```

(or if you prefer

```
[b+a]−a = b
```

)

Thanks to the commutative law, this is then the same as

```
[a+b]−a = b.
```

In other words, if you have a number that has been formed by addition, you can get either operand of the addition by subtracting the other operand.

Subtraction on the other hand is not commutative:

```
2−3 ≠ 3−2.
```

This means that although we have

```
[a−b]+b = a,
```

we don’t have

```
[a−b]+a = b.
```

To get `b` we can do a number of things: we can reverse the operands, and swap the operator:

```
a−[a−b] = b,
```

or we can negate the first operand and add:

```
−[a−b]+a = b,
```

or we can define a new operator, let’s call it *co-plus* ⊕, whose definition is

```
(a⊕b)=(−a)+b,
```

and then say that

```
[a−b]⊕a=b.
```

Note that both of these latter two require the definition of unary −, which we will talk about shortly.

Now I can understand if you are looking at this wide-eyed, wondering what I am going on about. So let me try and explain.

If I say that two people's combined ages are 95, then if I tell you one of the ages, you can work out the other without any more information. If I say one is 50, then you know the other is 45, and vice versa. In both cases you subtract the age from the sum of the ages. This is because addition is commutative.

However, if I had said that the *difference* in age between two people is 15 years and one of them is 45, you don't know enough, because you don't know if 45 is `a` or `b`. The other one could be either 60 or 30. So you need two complements for subtraction, one to get the left operand and the other to get the right operand.

Mystery Numbers

In the act of defining subtraction there is something surprising that has happened, that modern people might not even notice, but that must have upset early mathematicians. Namely, it has generated new sorts of number that we didn't start off with. All of a sudden we can calculate results like "0−5" which is "the number that adding five to gives zero" (or "the number that you would get if you subtracted 5 from 0").

But the problem is, how do you write such a number down? You can't write it with stripes, like we started off with, because they are a sort of "anti-number". It must have upset early mathematicians, because you now have numbers that apparently don't exist in the real world, but they seem to obey all the rules of regular, real-world numbers.

Well, as is so often seems to be the case with new types of numbers, they got given a very derogatory name. They could have chosen "mystical numbers" or "magical numbers" or "unreal numbers", but they chose "negative numbers". But they still had no way to write these numbers down.

Through the ages people have written negative numbers in different ways. Banks used to write them in red ink (which is why we (still) use the phrase “in the red”); annual reports of companies often use a number in brackets like “(123)” to mean a negative amount.

In fact, in mathematics we leave negative numbers uncalculated: “0−5” means “the number you would get if you subtracted 5 from 0”. Since it was always 0 that was used, it could be shortened to just “−5”; this is not strictly speaking a negative number, even if that is how we think of it nowadays: it is an operation on a positive number, an operation that we leave uncalculated, since it is good enough, or at least, the best we have.

Zero

A question that even now all mathematicians don't completely agree on is whether before defining subtraction we already had zero, or did the act of defining subtraction introduce zero?

Certainly early mathematicians didn't think of zero as a number (they even had difficulties with 1): if you didn't have any sheep, how could you say that you had a number of sheep?

Either way, mathematicians now agree that zero is a number, just not whether it was there right from the start.

Zero has a special relationship with addition and subtraction. It is called the *identity value* for addition and subtraction. This is shown by the identity law for addition and subtraction:

```
a+0 = a
a−0 = a
```

In other words `a` remains the same (it 'keeps its identity') if you add 0 to it, or subtract 0 from it.

Multiplication

Now we're going to go a level higher, and do almost exactly the same we just did with addition.

You can describe multiplication in terms of repeated addition:

```
a×3 = a+a+a.
```

The complement of multiplication is division, which we can define in the same way as we did for subtraction:

```
[a×b]÷b = a.
```

This means "find the value such that multiplying it by the second operand gives us the first operand". Just as with addition, multiplication is commutative, and so we can also derive

```
[a×b]÷a = b.
```

Similarly, just like subtraction, division is not commutative, since $2 \div 3 \neq 3 \div 2$. So again, although we have

```
[a÷b]×b = a,
```

we do not have

```
[a÷b]×a = b.
```

Again, we reverse the operands, and swap the operator:

```
a÷[a÷b] = b.
```

One

Multiplication and division also have an identity value, `1`, since

```
a×1 = a
```

and

```
a÷1 = a.
```

We've already discussed how `−a` is a shorthand for `0−a`, "the number you would get if you subtracted `a` from 0", and mentioned that `0` is the identity value for addition and subtraction.

It is therefore notable that there is no unary operator `÷a` to represent `1÷a`. However, if you introduce it, there are some rather pleasant symmetries that arise. For instance

```
a+(−b) = a−b
a×(÷b) = a÷b
```

and

```
a−(−b) = a+b
a÷(÷b) = a×b
```

and

```
−(−a)  = a
÷(÷a)  = a
```

Do you see the patterns?

There is something else interesting too, which you can see from this comparison:

```
−(a−b) = b−a
÷(a÷b) = b÷a
```

What you can see from this is that unary − and ÷ are both “commute” operators. They have the effect of commuting the operands of the thing they are applied to (without having to know what the values of those operands are).

You could say that while + and × are commutative, − and ÷ are commut*able*.

Earlier we said that to extract the left-hand operand from `(a−b)` we could use `(a−b)+b`, but to extract the right-hand operand we had to either use `a−(a−b)` or `−(a−b)+a`. This shows that from a conceptual point of view, the second of these has a certain charm: `(a−b)+b` extracts the left-hand operand `a`, and `−(a−b)+a` commutes the operands, giving `(b−a)+a`, and then extracts the left-hand operand of the result, giving `b`, the right-hand operand of the original.

So just as we defined the extraction of `b` from `a−b` as `−(a−b)+a`, we could also define the extraction of `b` from `(a÷b)` as `÷(a÷b)×a`.

Or we could define a new operator *co-times*⊗ defined as `(a⊗b)=(÷a)×b`, which commutes and extracts the left-hand operand, and use `(a÷b)⊗a=b`.

The ÷ unary operator will be used some more shortly.

More Mystery Numbers

Just as with subtraction, division introduces a new sort of number that we didn't start with.

We started with just positive numbers; thanks to subtraction, we got zero and negative numbers, but now thanks to division we have yet another sort of number. We can get numbers such as 2÷3, "the number that if you multiply by 3 gives you two". Although this is often written as $\frac{2}{3}$, we should recognise it as another example of an incompletely calculated number.

At least this branch of numbers got a friendlier name: the rational numbers. It is also worth pointing out that even though we didn't start out with these numbers, they still follow the rules of arithmetic for addition and subtraction.

The true definition of multiplication

To be honest, when I defined multiplication, I did a bit of hand waving, by just giving an example. I said

$$a\times3 = a+a+a$$

But this is not a true definition, just an example of what multiplication is. So let me now give the true (declarative) definition:

$$a\times(b+c) = a\times b + b\times c$$
$$a\times1 = a$$

What does this say? It says that you can split a multiplication up into a series of additions. (Again you might prefer the notation $a\times[b+c]$.)

For instance 3×4:

$$3\times4 = 3\times(2+2) = 3\times2 + 3\times2$$

$$3\times2 = 3\times(1+1) = 3\times1 + 3+1$$

So therefore

$$3\times4 = (3\times1 + 3+1) + (3\times1 + 3+1)$$

And since

$$3\times1 = 3$$

we have

$$3\times4 = 3 + 3 + 3 + 3.$$

You could do the same by starting out with

$$3\times4 = 3\times(3+1)$$

and you would still end up with

$$3\times4 = 3 + 3 + 3 + 3.$$

But how about

$$3\times0$$

though?

Well, we know that

$$3\times1 = 3$$

And we also know that

$$1+0 = 1$$

So therefore we can work out that

$$3\times(1+0) = 3\times1 + 3\times0$$

and since 1+0 = 1, we can also work out that

$$3\times1 = 3\times1 + 3\times0$$

and so we have to conclude that

$$3\times0 = 0.$$

Power

Now to go up yet one more level, using exactly the same patterns.

We can describe the operation of taking something to a power in terms of repeated multiplication:

$$a^3 = a \times a \times a.$$

Unfortunately, mathematics notation becomes odd at this level, and rather than using operator-style, like +, −, ×, and ÷, it starts using strange layout-style notations, as if there were something different going on.

Let's try and fix this. Here we will use a notation that is already used in some computer circles to represent taking the power, the ↑ operator, so that $a \uparrow 3$ means what classically is meant by a^3.

Of course, I've waved my hands above again, but this time I hope you can work out for yourself what the true declarative definition of power is: it uses the same pattern as multiplication.

If you do it, you will see that we can work out that just as

$$a \times 0 = 0$$

(which is the identity value for addition) we get

$$a \uparrow 0 = 1$$

(the identity value for multiplication).

The complement of power

Without thinking (yet) about what it means, we can also invent a complementary operation for ↑, which we will write as ↓; if ↑ is a higher-level form of multiplication, then ↓ is a higher-level form of division. Just as in the other cases, we define

```
(a↑b)↓b = a.
```

Unfortunately there is a difference at this level, since unlike + and ×, ↑ is not commutative: 2↑3 is 8 and 3↑2 is 9, which are clearly not equal, so we cannot just juggle operands as we did with addition and multiplication.

Instead we will invent another complementary operator to get `b`. Let's call it ⇓, so that we have

```
(a↑b)⇓a = b.
```

Well, let's now explain what these two operators mean: ↓ means "find the value that when taken to the power of the second operand gives us the first operand". For instance

```
8↓3 = 2,
```

since

```
2↑3 = 8.
```

In other words, it takes a root, so that a↓b means what is classically expressed with the, frankly bizarre, notation $^{b}\sqrt{a}$.

Similarly, ⇓ means "find the value that when the second operand is raised to that power, gives the first operand", for instance

```
8⇓2 = 3.
```

In other words, it takes the logarithm, so that a⇓b means what is classically, and equally bizarrely, written $\log_b a$.

(It can help, while learning to get used to these operators to note a slight visual similarity between √ and ↓, and to regard ⇓ as looking a bit like a log, while not forgetting that the operands are the other way round to how they are normally written.)

So now to properly deal with the two new operators: how do you extract the operands from them?

For ↓ we have

```
[a↓b]↑b = a,
```

and

```
a⇓[a↓b] = b.
```

For ⇓ we have

```
b↑[a⇓b] = a
```

and

```
a↓[a⇓b] = b.
```

The nice thing about the new notation is that it far more visually obviously expresses the relationships between power, root and log. So let us summarise the relationships, to more clearly show the patterns involved:

Operator	**a**	**b**
a+b a−b	(a+b)−b (a−b)+b	(a+b)−a a−(a−b) *or* −(a−b)+a *or* (a−b)⊕a
a×b a÷b	(a×b)÷b (a÷b)×b	(a×b)÷a a÷(a÷b) *or* ÷(a÷b)×a *or* (a÷b)⊗a
a↑b a↓b a⇓b	(a↑b)↓b (a↓b)↑b b↑(a⇓b)	(a↑b)⇓a a⇓(a↓b) a↓(a⇓b)

Just as when we defined the unary version of ÷, and we got some new laws that had the same patterns as with unary −, with these three new operators, we can observe some pleasant similarities with laws that we already know from arithmetic with the addition and multiplication operators. For instance, the basic operations:

Addition	**Multiplication**	**Power**
a+0 = a a−0 = a	a×1 = a a÷1 = a	a↑1 = a a↓1 = a
a−a = 0	a÷a = 1	a↓a = 1 a⇓a = 1
	a×0 = 0 0÷a = 0	a↑0 = 1 0↓a = 0 1⇓a = 0
	a÷0 = undefined	a↓0 = undefined a⇓0 = undefined a⇓1 = undefined

Similarly, there are several patterns with the unary operators:

Addition	**Multiplication**	**Power**
	$a\times(-b) = -(a\times b)$ $a\div(-b) = -(a\div b)$	$a\uparrow(-b) = \div(a\uparrow b)$ $a\downarrow(-b) = \div(a\downarrow b)$
$a+(-b) = a-b$ $a-(-b) = a+b$	$a\times(\div b) = a\div b$ $a\div(\div b) = a\times b$	$a\uparrow(\div b) = a\downarrow b$ $a\downarrow(\div b) = a\uparrow b$
$a+(0-a) = 0$	$a\times(1\div a) = 1$	$a\uparrow(1\downarrow a) = 1$

And finally, there are lots of patterns over the binary operators:

Addition	**Multiplication**	**Power**
	$a\times(b+c) = a\times b+a\times c$ $a\times(b-c) = a\times b-a\times c$	$a\uparrow(b+c) = a\uparrow b\times a\uparrow c$ $a\uparrow(b-c) = a\uparrow b\div a\uparrow c$
	$(a+b)\times c = a\times c+b\times c$ $(a-b)\times c = a\times c-b\times c$ $(a+b)\div c = a\div c+b\div c$ $(a-b)\div c = a\div c-b\div c$	$(a\times b)\uparrow c = a\uparrow c\times b\uparrow c$ $(a\div b)\uparrow c = a\uparrow c\div b\uparrow c$ $(a\times b)\downarrow c = a\downarrow c\times b\downarrow c$ $(a\times b)\Downarrow c = a\Downarrow c+b\Downarrow c$ $(a\div b)\downarrow c = a\downarrow c\div b\downarrow c$ $(a\div b)\Downarrow c = a\Downarrow c-b\Downarrow c$
	$(a\times b)\times c = a\times(b\times c)$ $(a\div b)\div c = a\div(b\times c)$	$(a\uparrow b)\uparrow c = a\uparrow(b\times c)$ $(a\downarrow b)\downarrow c = a\downarrow(b\times c)$
$(a+b)-c = (a-c)+b$ $(a-b)-c = (a-c)-b$	$(a\times b)\div c = (a\div c)\times b$ $(a\div b)\div c = (a\div c)\div b$	$(a\uparrow b)\downarrow c = (a\downarrow c)\uparrow b$ $(a\uparrow b)\Downarrow c = (a\Downarrow c)\times b$ $(a\downarrow b)\downarrow c = (a\downarrow c)\downarrow b$ $(a\downarrow b)\Downarrow c = (a\Downarrow c)\div b$
$a-b = (a-k)-(b-k)$	$a\div b = (a\div k)\div(b\div k)$	$a\Downarrow b = (a\Downarrow k)\div(b\Downarrow k)$
$a+(b-c) = a-(c-b)$	$a\times(b\div c) = a\div(c\div b)$	$a\uparrow(b\Downarrow c) = a\downarrow(c\Downarrow b)$
$a-b = -(b-a)$	$a\div b = \div(b\div a)$	$a\Downarrow b = \div(b\Downarrow a)$

Analysis

Looking at these similarities, there is one group that particularly sticks out:

$a+(-b) = a-b$	$a\times(\div b) = a\div b$	$a\uparrow(\div b) = a\downarrow b$
$a-(-b) = a+b$	$a\div(\div b) = a\times b$	$a\downarrow(\div b) = a\uparrow b$

This seems to suggest there should be a unary ↓ operator:

$a+(-b) = a-b$	$a\times(\div b) = a\div b$	$a\uparrow(\downarrow b) = a\downarrow b$
$a-(-b) = a+b$	$a\div(\div b) = a\times b$	$a\downarrow(\downarrow b) = a\uparrow b$

Although $\downarrow b$ means the same as $\div b$, the other unary operators are expressed in terms of their own level of operators, and so you can say that $\downarrow b$ means $b\uparrow -1$, which you can prove is the same as $b\downarrow -1$. This also has the nice property that $a\uparrow(\downarrow a) = 1$, just as $a\times(\div a)=1$ and $a+(-a)=0$.

It can also be applied to the last set of equations above:

$a-b = -(b-a)$	$a\div b = \div(b\div a)$	$a \Downarrow b = \downarrow(b \Downarrow a)$

Note its pleasant effect as a commuting operator for $\Downarrow$.

Yet More Mystery Numbers

Of course, I hope by now you are expecting us to come out of this level with more types of numbers than we began with, because you won't be disappointed: we get *two* new types.

For several centuries it has been known that a simple expression such as 2↓2 (the square root of 2) cannot be represented by a rational number.

This class of numbers got the derogatory name "irrational" (although the 9th century mathematician al-Khwārizmī called them *inaudible*, which is why we still call any unresolved irrational root a *surd*, Latin for *deaf*.)

These numbers also still work at the lower levels of multiplication, division, addition and subtraction.

The other new type of number introduced by the complement of power is the horribly misnamed *complex numbers,* which historically emerged as solutions of expressions such as (−2)↓2. These will be the topic of a later section.

Representations

How you represent numbers can have a great effect on what you can do with them.

For instance, the Romans may well have been able to add CCCLXI and CCXXI together, but few would have been able to multiply them together.

And yet nowadays, ask a schoolchild to add 361 and 221, or even multiply them together, and they will be able to do it for you.

In fact even in the middle ages, multiplication was still something you learned at university. It wasn’t until the 1600’s that modern numbers started to be introduced and used.

Our modern numbers come from Arab mathematicians, around the year 1000, who in turn got them from Indian mathematicians, from around the year 600.

Look at this piece of Arabic text, taken from the Wikipedia article in Arabic about the Second World War. Does anything about it strike you?

الحرب العالمية الثانية: هي نزاع دولي مدمر بدأ في الأول من سبتمبر 1939 في أوروبا وانتهى في الثاني من سبتمبر 1945، شاركت فيه الغالبية العظمى من دول العالم، في حلفين رئيسيين هما: قوات الحلفاء ودول المحور. وقد وضعت الدول الرئيسية كافة قدراتها العسكرية والاقتصادية والصناعية والعلمية في خدمة المجهود الحربي، وتعد الحرب العالمية الثانية من الحروب الشمولية، وأكثرها كلفة في تاريخ البشرية لاتساع بقعة الحرب وتعدد مسارح المعارك والجبهات فيها، حيث شارك فيها أكثر من 100 مليون جندي، وتسببت بمقتل ما بين 50 إلى 85 مليون شخص ما بين مدنيين وعسكريين، أي ما يعادل 2.5% من سكان العالم في تلك الفترة.

Well, of course they use what we call Arabic numerals. But the interesting thing is that Arabic is written from right to left, and yet the numbers are written in exactly the same way as we do, 1939, and not, as you might expect 9391.

How come? Well, when Western mathematicians imported our modern numerals from Arabic, they forgot to swap the order from right-to-left to left-to right.

Does this matter? A little.

Try adding two large numbers up in your head: add 1939 and 1945.

Do it now.

The answer if you got it right is 3884. But did you notice? You had to do all the addition for the whole number before you could start saying what the result is.

However, if numbers were the other way round, and I asked you to add the same two numbers, now written 9391 and 5491, you could speak the answer *while* doing the calculation:

    9+5 is **4** (carry 1)

    3+4 is 7, plus the carry **8**

    9+9 is **8**, carry 1

    1+1 is 2, plus the carry **3**

So the answer would be 4883 if they had swapped the direction on importing, and you would be able to do the calculation in your head and speak the answer while doing it. (Some computer programs on numbers would be simpler too.)

In fact for small numbers, we do speak our numbers that way round: we say FOURteen, SIXteen, SEVENteen. That is we say the second digit first.

Similarly, you probably know this children's rhyme:

> Sing a song of sixpence, a pocketful of rye.
> **Four and twenty** blackbirds baked in a pie.

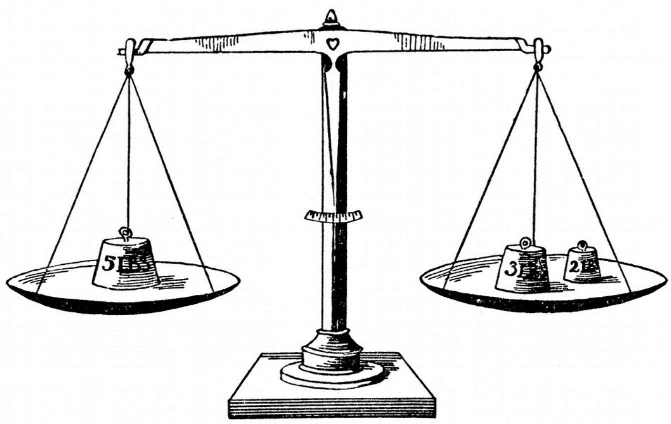
If you look at Shakespeare's plays, around two-thirds the usage of numbers is of the style "four and twenty", and the other third are of the modern style "twenty-four". This probably indicates that the style of speaking numbers was changing around Shakespeare's time.

So the conclusion is, the style of writing numbers we use now are much easier than Roman numerals, but they could be (a little bit) easier if they had been imported the right way round.

Representing Numbers

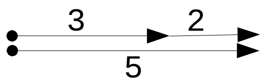
As was said at the beginning, all numbers are abstractions. Early mathematicians had problems accepting the concept of negative numbers, because they didn't seem to occur in real life. But in fact positive numbers don't 'occur' in real life either. As already pointed out, you can't point to "the number three".

However, we are not very good with abstractions in general. We like to think about them in some solid form, and so we represent numbers in some way or another in order to be able to talk about them. You can do that with three apples, you could do that with weights.



3+2=5 using weights

But since we are primarily visual, we typically represent numbers with lengths, drawing a line, and marking particular numbers on it. In mathematics, there is even the concept of “the real number line”.

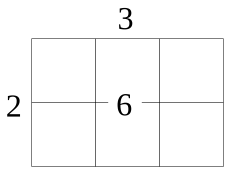


3+2 = 5 using lengths

However, there is nothing essential to this representation of numbers, and in fact a problem with concretisation of abstractions is that there is a risk of mixing up the two and thinking that the concretisation *is* the abstraction. A line is just one of many possible representations of numbers.

Sometimes we switch between representations too, often without realising it.

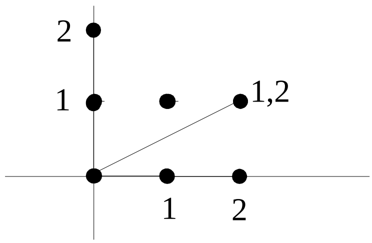
For instance, when explaining multiplication, we often draw a rectangle whose sides are of the length of the two numbers being multiplied, and explain that the area of the rectangle is the product of the two sides.



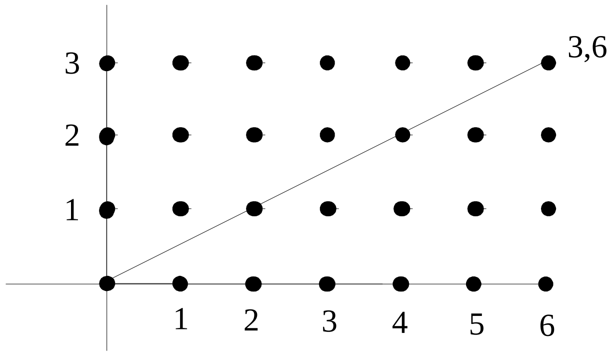
In so doing, we switch from a linear representation of numbers to a area-based representation.

Rational numbers, such as ¾, can be represented by a pair of integers (3, 4). One way of representing rational numbers is to create a two-dimensional field of dots, where in one dimension the integers are marked for the numerator, and in the other dimension for the denominator.

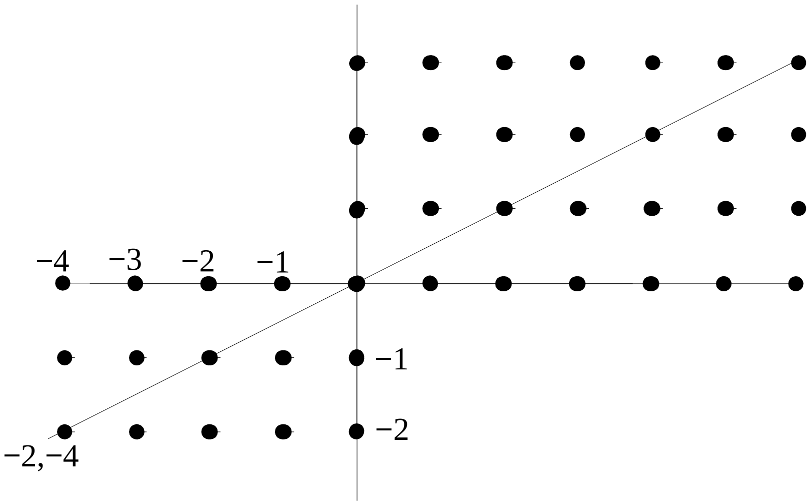
So drawing a line from the origin (0,0), through the point (1,2), represents the number ½.



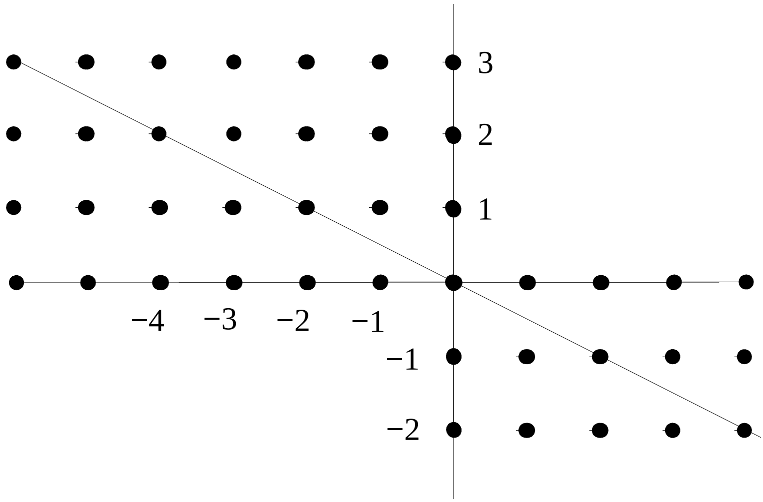
It’s the line that represents the number, and if you extend it, it passes through the points (2,4) and (3, 6) as well, demonstrating that these are equivalent rational numbers.



Furthermore, it passes in the other direction through (−1, −2) and (−2, −4) and so on, also in the process demonstrating that two minuses make a plus.

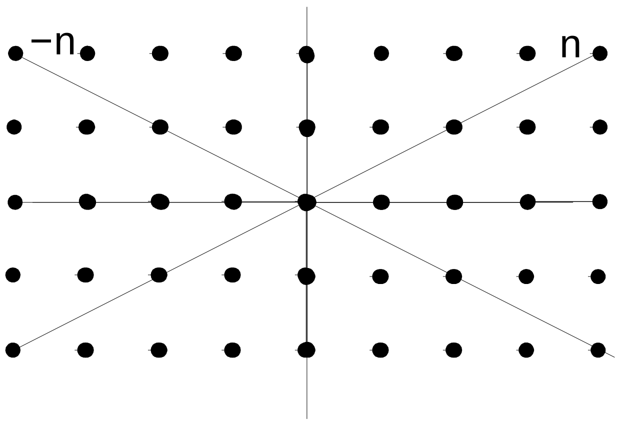


Similarly, a line from the origin to (−1, 2) represents the rational number −½, which likewise also passes through (−2, 4) as well as (1, −2) and (2, −4) , and so on.

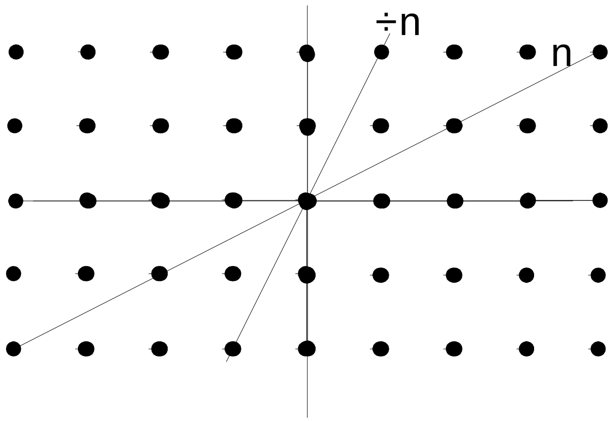


This representation has some other advantages. For instance, it also shows you that all points (0, x) represent zero, all points (x, 0) represent infinity (and actually allows you to represent infinity), and finally that (0,0) is indeterminate, since it represents all numbers, all lines passing through it.

It also has some other nice properties: the representation of the number −n is just the representation of n reflected by either the x or y axis.



Similarly, the inverse of n, ÷n, is n rotated 90 degrees (it doesn't matter in which direction), and negated (i.e. reflected on either axis).



So: if the angle of the line for the number n is a, then

```
−n = 180° − a
```

and

```
÷n = 90° − a.
```

This representation can also be used to describe how the line representing a non-rational, such as 2↓2 (the square root of 2), does not pass through any integral point at all, no matter how far you extend it, because it is not a rational number.

Why 2↓2 isn't a rational number

If 2↓2 is a rational number then it can be represented as

a÷b

for some a and b, reduced to their lowest form.

Note that when you have a rational number like that, that while a and b can both be odd numbers, they can't both be even, because if they are both even, you can divide them both by 2 to get a yet lower form. For instance

4÷6

can be reduced further to

2÷3

Reduced to their lowest form, they can both be odd, or one can be even, but they can't both be even.

So, if it is true that the square root of 2 can be represented by $a \div b$, then we have this:

$$2 \downarrow 2 = a \div b$$

Square both sides:

$$2 = (a \div b) \uparrow 2$$

Expand the brackets:

$$2 = a \uparrow 2 \div b \uparrow 2$$

Multiply both sides by $b \uparrow 2$:

$$2 \times b \uparrow 2 = a \uparrow 2$$

This means that a squared is an even number, and therefore a must be even too (an odd number squared is always odd).

Since a is even, b *must* be odd. Let’s try and find b.

Since a is even, then it is a multiple of some other number c:

$$a = 2\times c$$

So since we had

$$2\times b\uparrow 2 = a\uparrow 2$$

Substituting for a gives us

$$2\times b\uparrow 2 = (2\times c)\uparrow 2$$

Expand the brackets:

$$2\times b\uparrow 2 = 4\times c\uparrow 2$$

Divide both sides by 2:

$$b\uparrow 2 = 2\times c\uparrow 2$$

So $b\uparrow 2$ is even, which means using the same arguments as above, that b has to be even. So that means it is impossible to find an odd number that is the second operand of $a\div b$.

Since we can’t do it, it can’t exist. Therefore there are no integers a and b such that

$$(a\div b)\uparrow 2 = 2.$$

The square root of two is not a rational number.

In fact, it really shouldn't be so surprising.

If we go down one level, there are rational numbers

$$a \div b$$

that are not integers, just as there are roots

$$a \downarrow b$$

that are not rationals.

Why is there no ⇑ operator?

Since there is a ⇓ operator, why don't define the opposite of it as

```
(a⇓b)⇑b = a
```

just like

```
(a↓b)↑b
```

?

The answer is that we could do that, but it doesn't give us any extra functionality, since to get a from (a⇓b) we can use

```
b↑(a⇓b)
```

just as we don't need a new operator

```
(a−b)⊕a = b
```

to get b, since we can just use

```
a-(a-b) = b.
```

That is to say, that if we defined ⇑ like that we would have

```
a⇑b = b↑a
```

i.e., just the same operator, but with its operands the other way round.

Naming

How should we name the new operators? Again, it would be good to look to see if we can use consistency as a guide. So how are the existing operators named:

Operation	Written	Spoken
Addition	$a+b$	a plus b
Subtraction	$a-b$	a minus b
Multiplication	$a\times b$	a times b, or a multiplied by b
Division	$a\div b$	a over b, or a divided by b
Raising to the power	a^b	a to the power (of) b
Taking the root	$^b\sqrt{a}$	b-th root of a
Finding the logarithm	$\log_b a$	log to the base b of a

For addition and subtraction, the phrase just uses the names of the operators, neither left nor right operand playing any special role; for multiplication, a sort of adjectival form is used (“How many times have I told you?” “Four times”) to describe the process of multiplication, with the right hand side being the subject, or a form where the left hand side is the subject being acted upon by the right hand side; for division either a description of the layout is used (‘over’), or again the left hand side is the subject, which is acted upon by the right hand side; and for power, the left hand side is again the subject having something done to it.

So, from a consistency point of view, it's a mess, with little to draw from. I am inclined to use 'up' and 'down' for ↑ and ↓, modelled on plus and minus, although I sometimes pronounce something like "2↓3" as "two rooted three", when I want to emphasise the relationship to the traditional naming, using the "left hand side acted upon by the right hand side" model. That leaves us with ⇓, which I am inclined to pronounce as "logged" or "based", again using the left hand side as subject.

Examples

The purpose of this exercise was to develop, through consistency, a notation that is easier and more obvious to use than the classical notations. So let's put it to the test it with a few examples.

But first a word about solving equations.

If I say I bought two apples for 30 cents, almost without thinking you will know that each apple cost 15 cents. How? Well, calling the price of one apple `a`, we have

```
a×2 = 30
```

and we want to isolate a, which means we want to move everything to one side of the equals sign, except the a.

How do we do that? By applying the same operation to each side (so that the equality remains true), and then simplifying one or both sides.

In this case we divide both sides by 2:

```
(a×2)÷2 = 30÷2.
```

Simplify, and we're done:

```
a = 15.
```

But you should realise that there are some steps we have missed out, because they are ‘obvious’. Let me show you them, plus at each step the rule that is used:

```
a×2 = 30
(a×2)÷2 = 30÷2   {divide both by 2}
a×(2÷2) = 30÷2   {(a×b)÷c = a×(b÷c)}
a×(1) = 30÷2     {a÷a = 1}
a = 30÷2         {a×1 = a}
a = 15           {calculate}
```

As you can see, doing all the steps, although straightforward, is also tedious. So in the following, when we are isolating a value, we will often leave out the obvious steps with the word ‘simplify’.

Interest

If you save an amount of money, m, at an interest rate of 3%, then it means that at the end of the year

```
m×(3÷100)
```

gets added to your account, so you then have

```
m+m×0.03.
```

Factoring out m, this is the same as

```
m×(1+0.03),
```

which is

```
m×1.03.
```

In other words, each year your money gets multiplied by 1.03. So at the end of one year, you will have

```
m×1.03,
```

at the end of two years

```
m×1.03×1.03,
```

at the end of three years

```
m×1.03×1.03×1.03,
```

and at the end of n years,

```
m×1.03↑n.
```

So the general formula is:

```
result = m×(1 + rate÷100)↑years.
```

Let’s simplify this slightly and give the expression 1 + rate÷100 the name ‘r’, so that for 3% interest, r is 1.03. So our formula is

```
result = m×r↑years.
```

How much would you have to put in the bank at 3% to ensure that in five years you have 1000?

We have to isolate m. Take the equation,

```
result = m×r↑years
```

and divide both sides by r↑years:

```
result÷(r↑years) =
m×r↑years÷(r↑years)
```

Simplify the right-hand side:

```
result÷(r↑years) = m
```

So fill in values and calculate, and we get

```
1000÷1.03↑5 = 862.61
```

What interest rate would you have to have in order to double your money in ten years?

We have to isolate r. Take the initial equation

    result = m×r↑years

and divide both sides by m:

result**÷m** = m**÷m**×r↑years

Eliminate the m÷m×:

result÷m = **r↑years**

We want to isolate r, so take a root of both sides:

(result÷m)↓**years** =
(r↑years)↓**years**

replace (r↑years)↓years by r:

(result÷m)↓years = **r**

Since we want result to be twice m, and over ten years, we have to calculate:

    2↓10=r

which is approximately 1.07177 or in other words 7.177%.

How many years do you have to save at 3% to double your money?

We want to isolate years. Take the initial equation:

result = m×r↑years.

Divide both sides by m and simplify:

result**÷m** = r↑years

We want to isolate ‘years’, so take the log of both sides:

(result÷m)**⇓r** = (r↑years)**⇓r**

Simplify the right-hand side:

(result÷m)⇓r = **years**

m=1, result=2, r=1.03, which gives us:

2⇓1.03 = years

which gives us 23.45 years.

Computer Speeds

In 1965, Gordon Moore predicted that the density of components on integrated circuits was going to double every year at constant price, for at least ten years. Ten years later, he re-analysed the data, and increased the time to 18 months per doubling. Since then, his prophecy has held up fairly well, under the name "Moore's Law" (even though it isn't really a law, in any meaning of the word).

While he didn't actually predict that computers would get twice as fast per 18 months, that has been pretty much the result.

If computers get twice as fast every 18 months, then we can represent the relative speeds by:

```
speed = 2↑(months÷18)
```

What is the annual growth?

```
2↑(12÷18)
```

which is 1.59, in other words a 59% annual growth. The monthly growth is 2↑(1÷18), which is 1.04, or in other words, 4%.

If only banks offered that sort of interest...

How many months before you can buy a computer that is 10 times faster?

We want to isolate months. Again, start with the speed equation:

speed = 2↑(months÷18)

Take the log of both sides

speed⇓**2** = (2↑(months÷18))⇓**2**

Simplify the right-hand side:

speed⇓2 = **months÷18**

Multiply both sides by 18:

speed⇓2**×18** = months

For speed=10 we have:

10⇓2×18 = months

which is just under 60 months, or 5 years.

For a speed gain of 100, we calculate

```
100⇓2×18 = months
```

which is just under 120 months, or 10 years.

But really, we didn't need to calculate that because we already knew from the table of equivalences above that

```
(a↑b)⇓c = (a⇓c)×b
```

and since 100=10↑2, were were calculating

```
(10↑2)⇓2
```

which is the same as

```
(10⇓2)×2
```

Since we had just calculated 10⇓2, the answer had to be twice that answer.

Put another way: we knew that it takes 5 years to get a speed increase of 10; after another 5 years we would have got another speed increase of 10, and 10×10=100 times speed increase, and 5+5= 10 years.

Networks

Where I work was the first internet connection in Europe on the open (non-military) internet. In 1988 the first connection from Europe to the United States was in the office next to mine, and all of Europe was connected to all of the USA at the blisteringly fast speed of 64k bits/second. (Nowadays a mobile phone is typically 1000 times faster than that). A year later the speed doubled to 128 k bits/second, and we rejoiced.

But in fact, even better than the speed increase of computers, network bandwidth doubles per year. As of this writing, where I work is the second fastest internet node in the world, and peaks at 4.5T bits/second: it has more or less doubled per year ever since 1988.

And the same is true for home connections. They also average out at doubling per year at constant cost.

So, if you now have a 20Mbps connection, how long before you could expect 1Gbps? Well, 1G is 1024M:

```
20×2↑y = 1024
```

We want to isolate y. Divide both sides by 20

```
2↑y = 51.2
```

Take the log of both sides

```
(2↑y)⇓2 = 51.2⇓2
```

Simplify the left-hand side:

$$y = 51.2 \Downarrow 2$$

Which is about 5.68 years.

I said that it has “more or less doubled per year”. But what was the yearly growth really? Let’s call it g. As I write, it has been increasing for 28 years; 1 Kb is 1024 bits, and 1 Tb is $1024 \uparrow 4$ bits.

$$64 \times 1024 \times g \uparrow 28 = 4.5 \times 1024 \uparrow 4$$

We want to isolate g: divide by 64×1024:

$g \uparrow 28 = (4.5 \times 1024 \uparrow 4) \div (64 \times 1024)$

Simplifying

$$g \uparrow 28 = (4.5 \times 1024 \uparrow 3) \div 64$$

$$g \uparrow 28 = 4.5 \times 16 \times 1024 \uparrow 2$$

$$g \uparrow 28 = 72 \times 1024 \uparrow 2$$

Take the root

$$(g \uparrow 28) \downarrow 28 = (72 \times 1024 \uparrow 2) \downarrow 28$$

Simplify and calculate:

$$g = 1.91$$

So it hasn’t grown at 100% per year, but 91%.

Population Growth

Let’s assume that world population growth is exponential, which means that over some fixed period of y years it doubles. Let’s call the yearly growth g; then what this means is that

$$g\uparrow y = 2$$

So, if we want to know what the yearly growth g is, we take the root of both sides:

$$(g\uparrow y)\downarrow y = 2\downarrow y$$

and simplify:

$$g = 2\downarrow y$$

which we can also write as $2\uparrow\div y$.

So, if each year the growth is $2\uparrow\div y$, then over a period of n years the growth is $2\uparrow(n\div y)$. The world grew from 6 billion to 7 billion in the 13 years up to 2012, so we have

$$6\times 2\uparrow(13\div y) = 7$$

What then is y, the doubling period? Divide both sides by 6:

$$2\uparrow(13\div y) = 7\div 6$$

Take the log of both sides:

$$(2\uparrow(13\div y))\Downarrow 2 = (7\div 6)\Downarrow 2$$

Simplify:

```
13÷y = (7÷6)⇓2
```

Divide both sides by 13:

```
÷y = ((7÷6)⇓2)÷13
```

Invert both sides:

```
÷÷y = ÷(((7÷6)⇓2)÷13)
```

Simplify

```
y = 13÷(7÷6)⇓2
```

Which you may also transform to this if you wish:

```
y = 13×2⇓(7÷6)
```

Which gives a doubling time of 58.45 years. So all other things being equal, we could expect a population of 14 billion in around 2070.

When might the population enter double figures?

```
7×2↑(x÷y) = 10
```

Divide by 7

```
2↑(x÷y) = 10÷7
```

Take the log:

```
x÷y = (10÷7)⇓2
```

Multiply by y

```
x = y×((10÷7)⇓2)
```

Which is about 30 years after 2012.

Why there is no perfection

Imagine you want to buy a folding bike. What do you want from such a bike? Well, for instance, that it is light, strong, cheap – or at least reasonably priced, easy to fold, quick to fold, small when it is folded, comfortable to ride... there are probably a dozen such properties that you want a folding bike to fulfil.

Well, let’s just take three of them for now: light, strong, and cheap. Unfortunately, search as you may, you won’t be able to find a folding bike that matches all three: you can find strong and light, but not cheap; you can find cheap and strong, but not light; you can find cheap and light, but not strong. Just those three constraints are not satisfiable. You *could* find strong, cheap and light, but then it wouldn’t be a bike: being a bike and being able to fold it are two constraints that are non-negotiable – it *must* be a bike, and it *must* be foldable.

So in other words, you will have to relax at least one of your negotiable constraints. You could decide not to go on holiday next year and use that extra money to buy a non-cheap bike. Or you could decide to put up with a heavy bike, or you could decide to be careful, and go with a not-strong bike. In any case, you have to put up with non-perfection.

And this is why you shouldn't expect to find a perfect partner either. What might you look for in a partner? Good looking, healthy, fit, financially secure, amusing, good conversationalist, good natured, good at cooking, musical, good in bed, ... There are lots of constraints, and they will be different for different people.

Then there are probably a few that you wouldn't think of mentioning, maybe because they are non-negotiable, like: of the opposite sex/same sex depending on your preference, in an age group not too different from your own, speaks a language that you also speak, interested in you, and so on, and so on.

The problem is, as you add each constraint, the pool of potential perfect partners gets smaller and smaller, and then you have to add the constraint that it is someone you will somehow actually get to meet in a social context... This is why you should prepare yourself for not meeting the *perfect* partner: you will need to relax some of your constraints to find someone who is at least satisfactory.

Which brings us, funnily enough, to musical tuning.

You probably know that a note in music is caused by air vibrating, at a different frequency for different notes. For instance, internationally it has been agreed that the note A above middle C has a frequency of 440Hz. Not for any particularly good reason (and a slightly better choice would be 430.5Hz, but not so much better to make it worth retuning all the instruments in the world) but 440Hz is good enough.

It turns out that our ears like combinations of frequencies that are related in certain simple ways, as small ratios of frequencies. For instance, a note played an octave higher is just twice the frequency, that is to say they have a ratio of 2:1. So the A one octave higher would be 880Hz. We hear a note an octave higher as almost the same note, only higher (and incidentally, it is also just by convention that we call it ‘higher’, a longer organ pipe plays a deeper note, so we could have called them the other way round).

The next-most simple ratio of frequencies is 3:2. This is classically the relationship between the notes called a "perfect fifth", between the first note on the major scale, and the fifth (which are seven semitones apart). So for instance, a fifth from A is the note E, so E would have a frequency of 440 × 3÷2, which is 660Hz. And the combination of an A and an E sounds nice to the human ear.

Of course, we want to be able to play a perfect fifth from any note, not just A, so the perfect 5th from E, which is B, would have a frequency of 660 × 3÷2, which is 990 Hz. And so we can build up a whole octave based on the premise that from every note we can also play the perfect fifth.

Since 990 is above 880, we can halve the number to take it down into the octave we are building up, between 440Hz and 880Hz. So the B is at 495Hz. The fifth above that is F# which will have a frequency of 742.5, and so we can step through C#, G#, D#, A#, F, C, G, up to D, which has a frequency of 594.67Hz. The fifth above that takes us back to A, which will have a frequency of 594.67 × 3÷2, which is ... 892 Hz? But didn't we say that A would be 880Hz?

Yes we did.

In other words, it is *impossible* to make an octave of notes where you can play a perfect 5th from every other note. We have to relax one of our constraints.

Well, how about if we take the next most simple ratio, the perfect fourth, which has a ration of 4:3? Can we build an octave out of that, starting with A=440, and its perfect fourth D being 440×4÷3 = 586.7 and so on? I'm afraid that the answer to that is also no: you end up with an octave A that has a frequency of 868.1Hz, instead of the required 880Hz.

So what is to be done?

It is clear hopefully from the above that the octaves are non-negotiable. Any solution has to have the octave of any note as twice the frequency, otherwise you would get awful dissonance.

So what we can try to do is divide the octave up into 12 equal steps, equal in the sense that each semitone has the same frequency ratio with its neighbour. But what is that ratio?

Let’s call the ratio r.

Starting from A, the calculation 440×r should give us A#. Then 440×r×r should give us B, and 440×r×r×r should give us C, and so on all the way up to the next A:

$$440\times r\times r\times r\times r\times r\times r\times r\times r\times r\times r\times r\times r = 880$$

(that’s twelve r’s). Writing this another way:

$$440\times r\uparrow 12 = 880$$

or

$$r\uparrow 12 = 880\div 440$$

or

$$r\uparrow 12 = 2$$

or in other words

$$r = 2\downarrow 12$$

Well, we know how to calculate that: r is just under 1.06.

So if we calculate the resultant octave, it looks like this:

A	A#	B	C	C#	D	D#	E	F	F#	G	G#	A
440	466.2	493.9	523.3	554.4	587.3	622.3	659.3	698.5	740	784	830.6	880
					586.7		660					

Underneath E I have shown what we would ideally have for a perfect fifth from A, and under D, a perfect fourth. What you can see is that the difference is very small, less than 1Hz in both cases. So small in fact that since we *want* to hear the right tuning, we think it is properly tuned (a lack of dissonance due to cognitive dissonance).

So, the conclusion is, don’t expect perfection, but if you relax some of your requirements you might just find something that so nearly matches that you can’t tell the difference.

Notes

So, the note A above middle C has a frequency of 440Hz, and each of the 12 semitones in an octave are separated by a factor of 2↑÷12. So what is the frequency of middle C? C is 9 semitones lower, so the frequency is

```
c=440×2↑(−9÷12)
```

which we know we can also represent as

```
c=440÷2↑(9÷12)
```

If we want, we can reduce it further to

```
c=440÷2↑(3÷4)
```

but either way, it calculates to 261.626Hz

Which note is closest to 512Hz?

```
512=440×2↑(n÷12)
```

Divide by 440

```
512÷440=2↑(n÷12)
```

Take the log

```
(512÷440)⇓2=(2↑(n÷12))⇓2
```

Simplify

```
(512÷440)⇓2=n÷12
```

Multiply by 12

```
((512÷440)⇓2)×12=n
```

Which is 2.62 (semitones higher than A). That means that the nearest note to 512Hz is C, which has a frequency of

```
440×2↑(3÷12)
```

which is 523.25Hz (not surprisingly twice that of middle C).

If C were set at 512Hz, what frequency would A be?

    512÷2↑(3÷12)

which can also be written:

    512÷2↓(12÷3)

which is

    512÷2↓4

or

    512×2↓−4

any of which give 430.54Hz.

A frequency of 512 is 2↑9, and since octaves are constant doublings, you can now see why 430.54Hz could be seen as a good frequency for A rather than 440. It is said that Mozart used to tune his piano to A430.

Earthquakes

The Richter scale of magnitudes of earthquakes represents the amplitude (height) of the seismic waves. For a magnitude of m:

```
amplitude = 10↑m
```

The energy released on the other hand is the 1.5th power of the amplitude:

```
energy = (10↑m)↑1.5
```

which, because of the rule

```
(a↑b)↑c = a↑(b×c)
```

can also be written

```
energy = 10↑(m×1.5).
```

This grows really fast: an earthquake of magnitude 4 releases 1 million units of energy; magnitude 6 releases 1000 times that.

(Actually, the Richter scale itself is no longer used, but a new scale measured in a different way is used, but has been tuned to match the Richter scale as closely as possible.)

So what is the difference in magnitude between two earthquakes, where the second releases twice as much energy as the first?

$$e_1 = 10\uparrow(m_1 \times 1.5)$$

$$e_2 = 10\uparrow(m_2 \times 1.5)$$

$$e_2 \div e_1 = 2$$

Substitute in:

$$10\uparrow(m_2 \times 1.5) \div 10\uparrow(m_1 \times 1.5) = 2$$

Use the rule $a\uparrow b \div a\uparrow c = a\uparrow(b-c)$:

$$10\uparrow(m_2 \times 1.5 - m_1 \times 1.5) = 2$$

Take the 10 log of both sides:

$$(10\uparrow(m_2 \times 1.5 - m_1 \times 1.5))\Downarrow 10 = 2\Downarrow 10$$

Simplify:

$$m_2 \times 1.5 - m_1 \times 1.5 = 2\Downarrow 10$$

Factor out the 1.5:

$$(m_2 - m_1)\times 1.5 = 2\Downarrow 10$$

Divide both sides by 1.5:

$$m_2 - m_1 = 2\Downarrow 10 \div 1.5$$

Which is marginally above 0.2. In other words, an earthquake of magnitude 7.4 is about twice as powerful as an earthquake of magnitude 7.2.

Personally I was amazed by this. Until I worked it out, I assumed, as I am sure many people do, that magnitude 7.2 and 7.4 earthquakes were quite similar, when in fact they aren’t at all. The 7.4 is twice as destructive!

How could we fix this?

Well, a simple fix would be to introduce a new scale by just multiplying the magnitudes by 5. Then a magnitude 7.2 earthquake would become a scale 36 earthquake and a magnitude 7.4 earthquake would become a scale 37 earthquake. People could then understand better that the next number in the scale is twice as powerful, while still having a close enough relationship to the magnitude that it would be easy to convert from one to the other.

However, as I said above, twice as powerful is “marginally above 0.2”, not *exactly* 0.2, and within a few points on the new scale the rule would diverge. For instance, the difference between a scale 36 earthquake and a scale 41 earthquake would actually not be 32 times as powerful as you might expect, but only 31.6. This is because the magnitude scale is based on powers of ten, and 31.6 is 1000↓2.

So what other possibilities are there?

Well, we could define a scale purely based on powers of 2. We could start at the same point, so that magnitude 0 was scale 0, and then *define* the new scale so that each point higher was *exactly* twice as powerful. What would that look like?

Well, as I said earlier, the magnitudes are the measure of amplitude, and the energy released is

$$e = 10\uparrow(m\times1.5)$$

The new scale would start at the same value for 0, which would be energy

$$10\uparrow(0\times1.5) = 1$$

and then double at each step:

$$e = 2\uparrow s$$

So putting these together, we have

$$10\uparrow(m\times1.5) = 2\uparrow s$$

Solving for m, we get:

$$m\times1.5 = (2\uparrow s)\Downarrow10$$

$$m = ((2\uparrow s)\Downarrow10)\div1.5$$

and solving for s we get

$$10\uparrow(m\times1.5) = 2\uparrow s$$

$$(10\uparrow(m\times1.5))\Downarrow2 = s$$

Now tabulating these, we would get the following:

Magnitude	Scale
1	4.98
2	9.97
3	14.95
4	19.93
5	24.91
6	29.90
7	34.88
8	39.86
9	44.85
10	49.82

(The largest recorded earthquake ever was magnitude 9.5, scale 47.34.)

Scale	Magnitude
5	1.00
10	2.01
15	3.01
20	4.01
25	5.02
30	6.02
35	7.02
40	8.03
45	9.03

As you can see, the difference between this and the rule “multiply the old scale by 5” is very small, so maybe it’s better to just stick with that.

You might actually like to compare the two equations for m and s, and admire their symmetry:

```
m = ((2↑s)⇓10)÷1.5
s = (10↑(m×1.5))⇓2
```

Arithmetic

For hundreds of years schoolchildren have been tortured with the mathematics of powers, roots and logarithms as if they were a completely different branch of thinking. However, if we change the notation to match that of simple arithmetic, we can do no more than conclude that they are actually just one more part of simple arithmetic, with similar rules and productions.

To illustrate this point, the Wikipedia article on logarithms ends the section on bases with the statement

> *"Given a number x and its logarithm $log_b(x)$ to an unknown base b, the base is given by:*
>
> *$b = x^{1/log_b(x)}$"*

Translating this into our notation we have:

    b=x↑÷(x⇓b)

which we know to transform to

    b=x↓(x⇓b)

To which we reply "Ha! That's the very definition of ⇓!".

As another example, the Wikipedia article on roots says

> *"Simplifying radical expressions involving nested radicals can be quite difficult. It is not immediately obvious for instance that √(3+2√2) = 1 + 2√2"*

So let’s prove it, and show how simple it now is! Convert to our notation:

(3+2×2↓2)↓2 = 1 + 2↓2

Square the right-hand side:

= (1 + 2↓2)↑2

Expand (a+b)↑2 to a↑2 + 2×a×b + b↑2:

= **1↑2 + 2×2↓2+ (2↓2)↑2**

Replace 1↑2 with 1, and (2↓2)↑2 with 2:

= **1** + 2×2↓2+ **2**

Reorder:

= **3**+2×2↓2

Take the square root:

= (3+2×2↓2)↓**2**

QED.

Do you really accept negative numbers as first-class citizens?

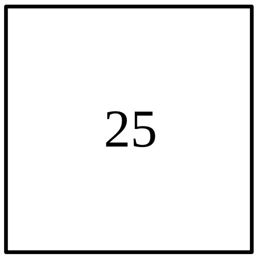
I’ll say it again: all numbers are abstractions. (There, I’ve said it three time, it must be true.)

However, it has taken humans a long time to realise this. For instance, for hundreds of years even mathematicians didn’t accept negative numbers, because they didn’t seem to correspond to anything in “the real world”. They were ‘fictitious’ or ‘false’ values, and solutions that gave negative numbers were ignored. You could count three sheep, but what could minus three sheep possibly mean? And addition was about making the result larger, ‘plus’ means ‘increase’, so how could a+b possibly be smaller than a, which would be the case with negative b? Pythagoras had no negative numbers for instance. Only recently has my bank been reporting debits as negative numbers, rather than positive amounts of debit.

Of course, we moderns accept negative numbers as full partners of the positive numbers, don’t we? We accept that a car travelling backwards can be regarded as travelling at a negative velocity. We accept that you can have a negative balance in your bank account. We know how to deal with the idea of negative temperatures.

So, now that that is out of the way, let me ask you a question.

A farmer has a square field, of area 25. How long are the sides of the field?



Well, I expect you answered 5, which is the right answer.

Well, at least, one of the right answers, because −5 would be a correct answer too.

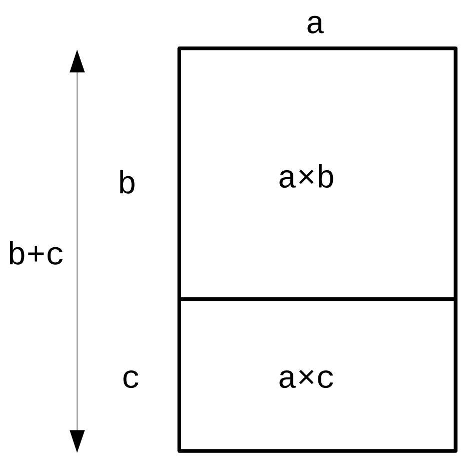
“But how can a field have a side of length −5?” I hear you ask. Now you know how early mathematicians felt about negative numbers.

So let me explain how a field can have a negative side.

Algebra and geometry often go hand in hand. We defined multiplication using

```
a×(b+c) = a×b + a×c
```

and there is a really good way to illustrate this equality, that shows the area of the whole rectangle is made up of the sum of the two small rectangles:



$a \times (b+c) = a \times b + a \times c$

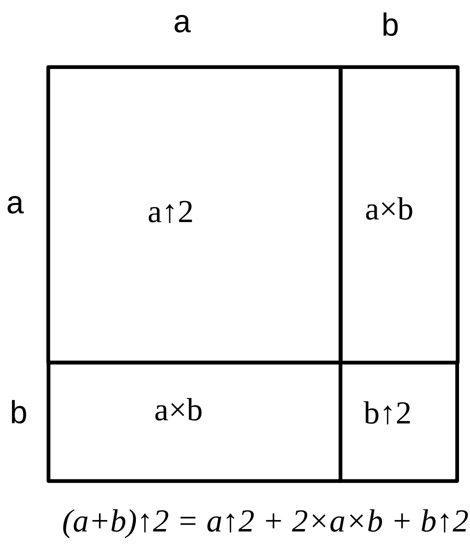
Another well-known equality (which can be derived from the one above) is

```
(a+b)↑2 = a↑2 + 2×a×b + b↑2
```

Let’s derive it:

```
(a+b)×(a+b)
= a×(a+b) + b×(a+b)
= a×a + a×b + b×a + b×b
= a↑2 + a×b + a×b + b↑2
= a↑2 + 2×a×b + b↑2
```

which likewise can be illustrated geometrically:



(a+b)↑2 = a↑2 + 2×a×b + b↑2

But an important aspect of such an equality is that *a*, *b*, and *c* can be *any* number. So while you can have *a*=3 and *b*=2:

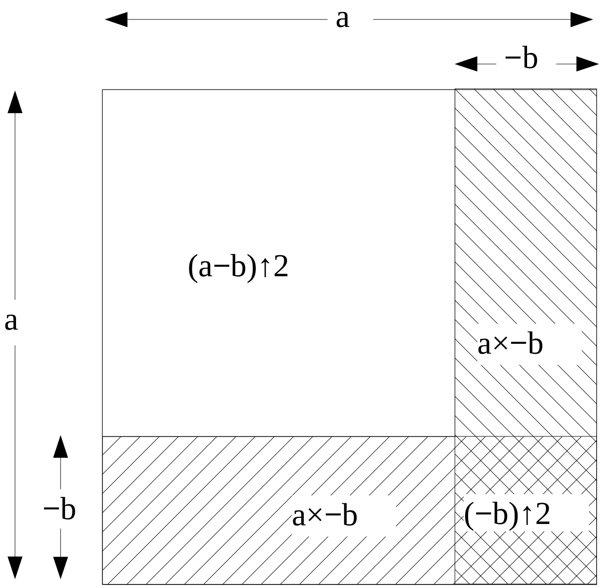
```
(3+2)↑2
= 3↑2 + 2×3×2 + 2↑2
= 9 + 12 + 4
= 25
```

you can also have *a*=3 and *b*=−2:

```
(3+(−2))↑2
= 3↑2 + 2×3×(−2) + (−2)↑2
= 9 + (−12) + 4
= 1
```

and the equality still holds.

But how does this version look in our geometric equivalent?

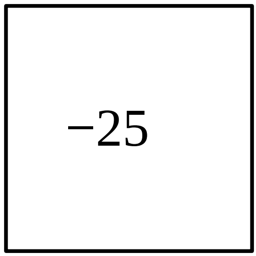


So the little square at the bottom right has sides of negative length −b, but a positive area (can you see why?) Even more interesting is that the two (overlapping) rectangles have negative area (a × −b)

So you can not only have shapes that have sides of negative length, but also shapes of negative area.

So let me ask another question to mirror the question asked at the beginning of this section.

Since we have now shown that it is possible to have shapes with negative area: what is the length of the sides of a square with area -25?

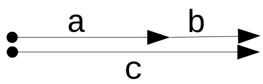


This is a question that we will come back to later.

Complete numbers

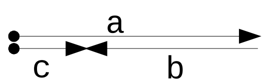
So even though society has for hundreds of years has problems with the concept of negative numbers, nowadays, we are happy to talk about negative temperatures, about negative growth, negative bank balances, and we accept that a negative speed indicates that the thing is travelling backwards.

One way of visualising addition of two numbers it to draw from zero a line of length the value of the first number, and then a line from the head of that line, of the length of the second number. The result is the length of the line from zero to the head of the second number:



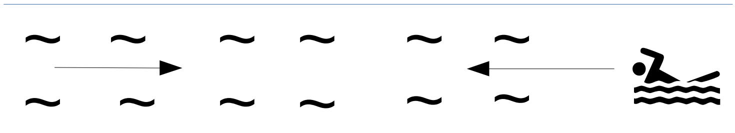
a+b=c

If any number is negative, then the arrow goes in the other direction:



a+b=c with b negative.

An observable application of this is someone swimming in a river. If someone is swimming at a fast 4 m/s in the same direction as the current of 3 m/s, then from the viewpoint of someone standing on the bank, they will be travelling at 7 m/s. Similarly, if they are swimming *against* the current, then from the viewpoint of the person on the bank, they will be travelling at 1 m/s.

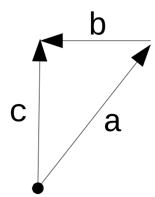


Swimming against the current

Generalising Numbers

But now we are going to free these numbers from the bounds of the number line, and let them point in *any* direction. A number is now its length *and* its direction. For reasons I will later explain, I will call these new, unrestricted, numbers *complete* numbers.

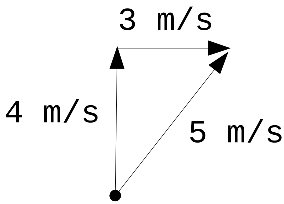
Addition is done in exactly the same way. You draw from zero a line of the length of the first number, pointing in the direction of the first number; from the head of that line you draw another line of the length of the second number, in the direction of the second number. Then the result is a line from zero to the head of the second number, and its direction is the resulting direction:



a+b=c with complete numbers.

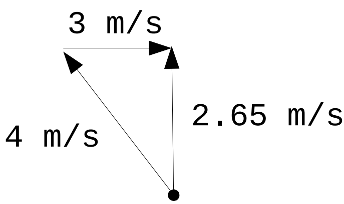
An immediate application of complete numbers is that you may now swim *across* the river, as well as up and down it.

So if you swim straight across the river at the speed of 4 m/s, and the river is flowing at 3 m/s, from the viewpoint of the person on the bank, you will end up swimming diagonally at a speed of 5 m/s:



It's the same addition rule, just more general. (4@90° + 3@0° = 5@53.1°).

Similarly, if you want to swim across the river so that from the bank it looks like you are swimming straight across, you have to swim slightly into the current:

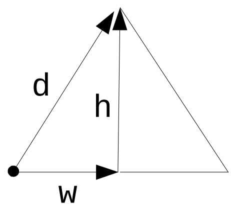


4@138.6°+3@0°=2.65@90°

Roof

But there are other applications of Complete addition.

If you know the height and width of a roof, you can easily calculate the length and angle of the roof covering:

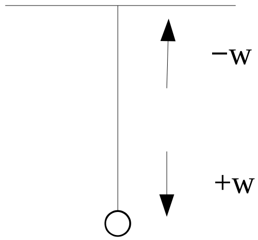


$$w+h = d$$

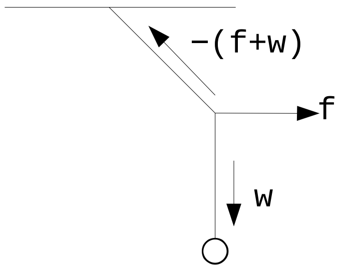
While we are looking at this example, it is worth mentioning that mathematicians would call the width a ‘real’ number, the height an ‘imaginary’ number, and the diagonal a ‘complex’ number (complex because it is both real and imaginary at the same time). I hope you can see why I think this is a misleading naming that should not be used, and all of them just be called ‘complete’ numbers.

Weights

If a weight hangs on a string from the ceiling, then the force in the string is just the same as the weight of the object, but in the other direction, so that they balance out:



However, if you attach another string to the middle of the first string, and pull on it, the force in the top half of the string is just the negative of the Complete sum of the two other forces:

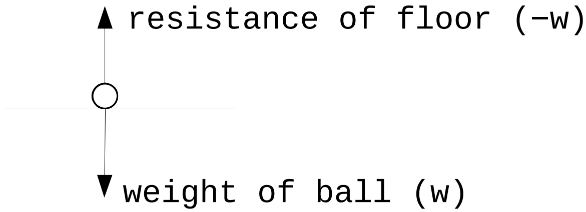


$$c = -(f+w)$$

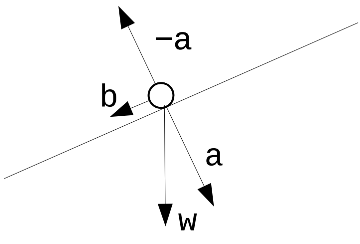
(‘Negative’ for complete numbers means ‘a half turn (180°) in the opposite direction’.)

Ball on sloping floor

If a ball is on a floor, then the floor offers the ball resistance exactly opposite to the force of the weight of the ball (otherwise the ball would sink, or crash, through the floor).



However, if the floor is sloping, the weight of the ball gets split into two forces, one which is at 90° to the floor, and one which is a force parallel to the floor, which pushes the ball down the slope. The complete sum of these two forces equals the force caused by the weight of the ball:

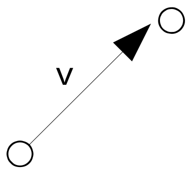


$$a+b=w$$

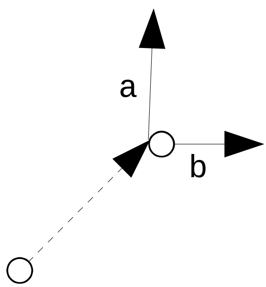
Billiard balls

If a moving billiard ball hits another, stationary, ball of the same weight, then the Complete sum of the velocities of the two balls after the collision is the velocity of the original ball.

Before:



after:



$$a+b=v$$

In fact, even if the original ball misses the other ball, it’s still true, because a+0 = a is still true with complete numbers.

Multiplication

OK, so all those examples use addition, but what does it mean to multiply two complete numbers together?

Well, as was already explained, multiplication has to obey the rule:

```
(a+b)×c = a×c + b×c
```

and while I shan't prove it here, the definition of multiplication that follows this rule is:

```
a@b × c@d = (a×c)@(b+d)
```

in other words, to multiply two complete numbers, you multiply the magnitudes, and add the angles.

A corollary of this is:

```
(a@b)↑2 = (a↑2)@(b×2)
```

and therefore

```
(a@b)↓2 = (a↓2)@(b÷2)
```

and in general:

```
(a@b)↑c = (a↑c)@(b×c)
```

and

```
(a@b)↓c = (a↓c)@(b÷c)
```

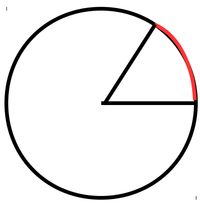
Angles

In the above examples, the angles of the Complete numbers have all been expressed in degrees, since that is the method of expressing angles that people are most at home with.

Why do we use 360° as the number of degrees in a circle? Possibly because it’s close to the number of days in the year; in any case it is a nice number to split up into pieces, because it is divisible by 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 18, and 20. Only 7, 11, 13, 14, 16, 17, and 19 are missing in that row.

However, rather than degrees, mathematicians prefer to use *radians* for measuring their angles (because it makes some laws simpler).

One radian is the angle such that the length of the arc drawn out by the angle is the same as the radius doing the drawing out; a sort of equilateral triangle, but with one curved side.



Since the circumference of a circle is $2\times\pi\times r$, there are $2\times\pi$ radians in a circle (that is $360° = 2\times\pi$ radians). Which means that one radian is equal to $360 \div (2\times\pi) = 57.3°$ (approximately).

However, for Complete numbers it is actually handier to talk in fractions of a turn τ, where 1τ is one turn, which is the same as 360° or 2π radians. The advantage is that it is easier to take the modulo, and more understandable to talk in terms of fractions of a turn.

For instance the square of

```
2@270°
```

is

```
(2×2)@(270°+270°)
= 4@540°
= 4@(540°−360°)
= 4@180°.
```

However, doing this in *turns*, we have

```
2@0.75τ × 2@0.75τ
= (2×2)@(0.75τ+0.75τ)
= 4@1.5τ
= 4@0.5τ,
```

since one and a half turns has the same effect as a half turn.

There is also the pleasant fact that if your take τ = 2π, you can still have your radians.

What is τ?

Just about everyone know what π (pi) is. If you measure the diameter of a circle, the circumference of the circle is π times longer. This is true whether the circle is the size of an orange, or the size of a planet.

If a wheel is 1 metre in diameter, then its circumference is π metres.

Almost everybody knows the approximate value of π too: somewhere near 3.14 (a little bit more, 3.141592...).

How can you use it?

Well, suppose you have a bicycle. A typical bike wheel is 68cm in diameter. That means that each time it revolves, you travel $68 \times \pi$ cm, about 214 cm.

When you turn your pedals, it drives that big cog wheel at the front. That cog is connected to the chain that connects to a similar, smaller, cog at the back. If the big cog has 52 teeth, then each complete turn of the pedals will push 52 links of the chain forward. If the cog at the back has 26 teeth, then the 52 links in the chain will drive the back cog (and the back wheel with it) $52 \div 26 = 2$ times round, which will cause the bike to travel $2 \times 68 \times \pi$ cm, which is 4.18 metres.

Presumably the diameter was chosen as the basis for π because it is easier to measure the diameter of a cylinder or wheel than the radius.

It wasn't until much later that mathematicians realised that it is better to use the radius than the circumference when talking about circles.

Unfortunately, when they made the change, they forgot to change π to go with it, which means that mathematics is now full of formulas that include $2\times\pi$ in them, such as the circumference ($2\times\pi\times r$).

But if we instead of using π, use τ (tau), with the value twice that of pi (in other words 6.283184…), then the circumference of a circle becomes $\tau\times r$, and the circumference of a quarter of a circle (90°) becomes $0.25\times\tau\times r$, and there are now τ radians in a whole circle, so that 180°, a half circle, becomes an angle of $0.5\times\tau$.

What about the square root of minus twenty five?

You may know that complex numbers (as complete numbers are traditionally called in mathematics) were born out a need to take the square root of negative numbers.

So to go back to the question that was asked earlier, what is the length of the sides of a field of area −25. Or to put it another way, what is the square root of −25?

Well, as we have now seen, a number like −25 is just another way of writing 25@180°, or rather, 25@0.5τ.

And, we know how to take the root of a Complete number:

```
(25@0.5τ)↓2 = 25↓2@(0.5τ÷2)
= 5@0.25τ
```

Nothing particularly special about that.

A note regarding the traditional treatment of imaginary numbers

The traditional way of handling imaginary numbers is to reduce them to real numbers multiplied by the unit of imaginariness, i, where $i = \sqrt{-1}$. Thus is $i \times i = -1$, and a complex number might look like `3+4i`.

If early mathematicians had adopted the same approach to their fictitious negative numbers, they would have observed that the unit of fictitiousness is f, where $f=0-1$, and noted that all fictitious numbers can be written as positive numbers times f, such as $5f$. All that you would have to remember is that $f \times f = 1$ to return us to the realms of the real world, so that for instance $5 \times 5f = 25f$, but $5f \times 5f = 25f^2 = 25$.

In other words, negative numbers would not have been given equal status with positive numbers, but treated as a poor sibling of them.

So is it with the traditional treatment of imaginary numbers.

A Note on Terminology

Although traditionally Complete numbers are called ‘Complex numbers’, which are made up of a combination of ‘Imaginary’ numbers and ‘Real’ numbers, they are neither imaginary nor complex, at least, no more imaginary than negative numbers, and no more complex than rational numbers. For that reason I call them something reminiscent of their old name, while trying to be fairer to them. It could be argued that these names have been used for so long now we should just accept them as they are, noting that they don’t describe what they refer to, and move on. However, the names are both off-putting for newcomers, and misleading for laypeople, so it is better for mathematics to rename them to something more representative.

Afterword

This book came from the combination of three different thought-streams: the first was while I was looking for good examples of mathematical functions that have two different inverses; the second was trying to understand complex numbers properly, and why they are treated so mystically in most maths books; and the third was while reviewing a son’s homework, and thinking “Why do they make something so simple so difficult?”

Hopefully, reading the text, it will all seem rather clear and straightforward. However, working out the new notations consumed *reams* of paper as I tried different options, and compared different patterns of formulas. The result may seem obvious, but it went through many variations as I experimented with different options.

As a colleague of mine who tries to design straightforward user interfaces once remarked: if you make it as easy to use as a coffee machine, they think of you as a plumber.