QUESTIONS PUPILS ASK:

WHY AREN'T SQUARE ROOTS ADDITIVE?

By **Colin Foster**

I've never had a student pose this question to me in precisely this technical kind of way. But I've certainly had many occasions in which a student seems to have assumed something like

$$\sqrt{a+b} = \sqrt{a} + \sqrt{b} \qquad (1)$$

and then been surprised to discover that it isn't true. "Why not?" might be a perfectly reasonable question for them to ask. But how would you answer it?

Substituting numbers

Perhaps the most obvious way to respond is to invite the student to try their equation with numbers. If algebra is 'generalised arithmetic', then a statement like (1) is really a claim about pairs of numbers (a, b) and how they behave. It makes sense to specialise and check it out with some example pairs of numbers.

In this case, finding a counterexample is extremely easy. Nearly *every* pair of numbers is a counterexample. Doing this, students may discover not only counterexamples like

$$\sqrt{1+2} \neq \sqrt{1} + \sqrt{2},$$

but they may also realise that the statement itself only makes sense if both a and b are non-negative. There could be some interesting discussion about whether substituting negative values of a and b *disproves* the statement or just indicates that we should have specified it more precisely by saying $a, b \geq 0$.

Through trial and error, students may discover that (1) is true when *both* a and b are zero. In fact, (1) is true when *either* a or b is zero, and this distinction might be worth thinking about and discussing.

We could prove that this solution is unique (i.e. that there are no other solutions) by squaring both sides of (1):

$$(\sqrt{a+b})^2 = (\sqrt{a} + \sqrt{b})^2 \qquad (2)$$

We have to be careful with this step, because squaring both sides can introduce spurious solutions. We would have got to the same place as (2) if we had started with (3), rather than (1):

$$-\sqrt{a+b} = \sqrt{a} + \sqrt{b} \qquad (3)$$

So, our solutions to (2) will *also* be solutions to (3). However, in this case, the left-hand side of (3) is never positive, and the right-hand side of (3) is never negative, so the only possible solutions to (3) will be $a = b = 0$. This means that we won't generate any additional solutions by squaring (1).

If we expand (2), we get

$$a + b = a + 2\sqrt{ab} + b,$$

from which it follows that $\sqrt{ab} = 0$. This means that $ab = 0$, meaning that either $a = 0$ or $b = 0$. There is some good reasoning involved in going through this and not getting confused and concluding $a = b = 0$, which is not the same. The argument is identical to the final step of solving a quadratic equation by factorisation, where we use this *zero-product property*.

When working on surds, it can be nice to contrast something like $\sqrt{3} + \sqrt{12} = \sqrt{15}$, which is false, with $\sqrt{3} + \sqrt{12} = \sqrt{27}$, which is true, and to have students invent examples like this (Foster, 2022).

This all seems like useful thinking. But does any of this address the 'why'? It's easy to fall into the trap of answering 'why' questions by demonstrating 'that', but not touching the 'why'. The student might say, "OK, I see that it *doesn't work* in general. But why doesn't it work?" Answers like "Square rooting isn't distributive over addition" are just re-stating the fact in more formal language, rather than explaining *why*.

Inequalities

When something *isn't* equal, we can always use the 'not equal' sign:

$$\sqrt{a+b} \neq \sqrt{a} + \sqrt{b}$$

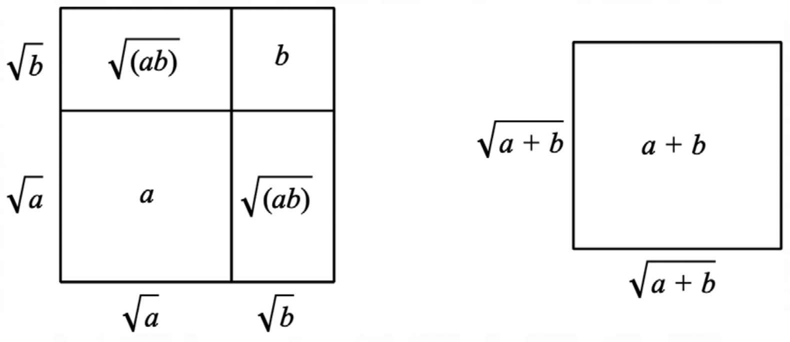
But it's even better if we can use an inequality sign, because then we're making a stronger statement. The student's experimentation so far may have suggested that

$$\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$$

with equality only when $a = 0$ or $b = 0$. The square root operation can be described as a *sub-additive function*. When you apply the function to a sum of several values, you never get more than if you apply the function to each of the values separately and add them up.

Can we see why this must be? The argument that we ran through above can do this. The right-hand side of (2) will always be greater than the left-hand side of (2), unless $a = 0$ or $b = 0$. So, $(\sqrt{a+b})^2 \leq (\sqrt{a}+\sqrt{b})^2$, with the equality only if either $a = 0$ or $b = 0$. A rectangular area diagram could help to show what's going on.

We can see in this diagram that the square on the right, with area $a + b$, must be smaller than the entire square on the left, with area $a + b + 2\sqrt{ab}$. They will only be equal if $\sqrt{ab} = 0$, meaning that either $a = 0$ or $b = 0$. Does this count as 'an explanation'?

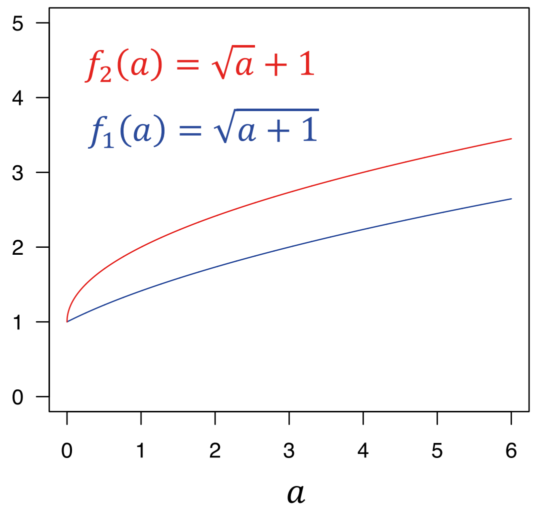


Graphing

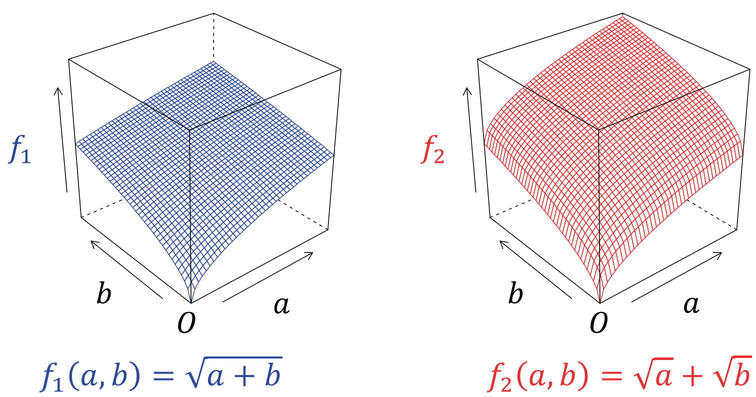
Perhaps a graph might help? Here we have two functions, both of two variables:

$$f_1(a,b) = \sqrt{a+b} \text{ and } f_2(a,b) = \sqrt{a} + \sqrt{b}.$$

This makes it challenging to draw a picture on a flat piece of paper. To make things easier, we could hold one variable constant (e.g. set $b = 1$) and vary the other, so comparing $f_1(a) = \sqrt{a+1}$ with $f_2(a) = \sqrt{a} + 1$.



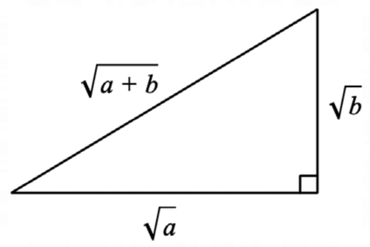
We see, to the right, the red curve (f_2) is never below the blue curve (f_1), and they coincide when $a = 0$. We could think about the translations needed to transform the red curve into the blue curve and vice versa, and this might provide additional insight into 'why'. Here are the surfaces in 3D.



Again, we see that the red surface (f_2) is never beneath the blue surface (f_1), and they match all the way along the bf and af planes, where either $a = 0$ or $b = 0$.

Pythagoras' Theorem

Where square roots are involved, Pythagoras' Theorem is rarely far away. We can take $\sqrt{a}$ and $\sqrt{b}$ to be the legs of a right-angled triangle, and this gives a hypotenuse of $\sqrt{a+b}$. By the triangle inequality (or 'common sense'), the hypotenuse can't be longer than the sum of the two legs, because the shortest distance across the hypotenuse must be the straight line that connects the two vertices that it joins.



So, $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, with equality only in the degenerate case, where either of the legs becomes of zero length, and the triangle collapses into a straight line.

Does this finally 'explain it'. I think I would say that it 'explains it' *in terms of Pythagoras' Theorem*. Perhaps all that an explanation in mathematics ever can be is a reduction to something else – ideally something perceived as more basic or already known or accepted. An explanation must always be 'in terms of something', and an explanation succeeds if it is given in terms of something that the student has already agreed to accept.

Other functions

Functional equations like $\sqrt{a+b} = \sqrt{a} + \sqrt{b}$ tend to appear more in advanced mathematics than in secondary school mathematics. But I think that there's plenty to explore at a more elementary level. A good task for students getting to grips with functions could be to explore *Cauchy's additive functional equation* – a grand name for $f(x+y) = f(x) + f(y)$. What kind of functions will satisfy this equation, and thus be termed *additive*?

The fact that the function $f(x) = \sqrt{x}$, $x \geq 0$ is concave downwards corresponds to it being *sub-additive*. A concave upwards function, such as $f(x) = x^2$, will instead be *super-additive*, meaning that $f(x+y) \geq f(x) + f(y)$. In this case, $(x+y)^2 \geq x^2 + y^2$, which students can show by expanding the left-hand side and revealing the lurking $2xy$ term.

The territory of sub-additive and super-additive functions covers a lot of common errors that students make by assuming that things are additive that aren't. For example,

$$(x+y)^2 = x^2 + y^2,$$

$$\sin(x+y) = \sin x + \sin y,$$

$$a^{x+y} = a^x + a^y.$$

Perhaps the language of additivity, sub-additivity and super-additivity, which I think is unfamiliar in schools, could be useful in discussing these things and making some of the differences more explicit.

There are lots of nice little 'theorems' to explore in this area, such as:

1. All non-increasing functions are sub-additive.
2. All non-decreasing functions are super-additive.
3. Functions which are both sub-additive and super-additive are additive.
4. Super-additivity is the 'dual' condition of sub-additivity, meaning that if f is sub-additive, then $-f$ is super-additive.
5. If an invertible function f is sub-additive, then f^{-1} is super-additive.

For more details, see Alsina & Nelsen (2009, pp. 119-120; 2010, pp. 229-230).

References

Alsina, C., & Nelsen, R. B. 2009 *When less is more: Visualizing basic inequalities* (No. 36), MAA.

Alsina, C., & Nelsen, R. B. 2010 *Charming proofs: A journey into elegant mathematics* (No. 42), MAA.

Foster, C. 2022 'Adding surds', *Teach Secondary*, 11(7), p.13. **www.foster77.co.uk/Foster,%20Teach%20Secondary,%20Adding%20surds.pdf**

Keywords: Additive, Functional equations, Functions, Misconceptions, Sub-additive, Super-additive, Square roots

Author: c@foster77.co.uk

website: **www.foster77.co.uk**