Communications

In Praise of the Minus Sign

AMIR ASGHARI

I have told this story orally several times since the day it happened in a middle school mathematics class about twenty years ago when I taught negative numbers to a group of thirteen years old students. Until recently, I always thought of it just as a good example of interactive concept creation in a mathematics class. However, following a recent research on the history of negative numbers (Asghari, 2019), the taken for granted side of the story, the negative numbers themselves, came into a new light. In particular, I learned some of the advantages of the common representation of negative numbers with a minus sign in front.

It might seem strange to talk about the advantages of such a familiar representation, in particular, when the Glossary of the Common Core State Standards for Mathematics (2010) defines an integer as follows:

> **Integer.** A number expressible in the form a or $-a$ for some whole number a. [1]

-3

So it seems it is defined as it is defined, and for that reason, we have to represent, say 'negative three' or 'minus three', by -3. But, we could use $^{-}3$ to distinguish between 'minus three' as in 'five minus three' and 'negative three' as a standalone number. Or we could even use different colours to distinguish positives from negatives, as the Chinese did more than two millennia ago, black for negatives, red for positives (Joseph, 2010). So why do we teach $-a$ instead of any other notation to represent negative numbers? After all, we know that it might be confusing for students to use the same symbol both for the operation of subtraction and a kind of number.

What if the negative numbers are represented in some other way? Here I first tell the story of some students who unintentionally (at least at the outset) experienced negative numbers represented in another way. Then, I compare the invented representation with the standard representation, and discuss the pros and cons of using the minus sign for representing negative numbers.

The students in my story were familiar with the notion of power (exponent) but not with negative numbers. The conversations are not verbatim, but in line with the real events.

The question

"Look at the table," instructed the teacher, the younger me.

					3^1	3^2	3^3			
					3	9	27			

"From left to right, the cells are multiplied by three; from right to left the cells are divided by three. Thus, we can fill the cells of the second row in both directions."

					3^1	3^2	3^3	3^4	3^5	
		$1/9$	$1/3$	1	3	9	27	81	243	

"We know that the cell above 81 is 3^4 and the cell above 243 is 3^5. But, what is the cell above 1, or above $1/3$? In other words, *what power of 3 goes with 1?*"

The first pattern

"Look," replied some of the students (perhaps half the class), "when you move from 3^2 to 3^1, the power becomes half. So, the next cell, should be $3^{1/2}$."

				$3^{1/2}$	3^1	3^2	3^3	3^4	3^5	
		$1/9$	$1/3$	1	3	9	27	81	243	

"It is true," they admitted, "that this does not hold when we move from 3^3 to 3^2 or from 3^4 to 3^3 (see, 3 is not half of 4). But, it does not matter. The cells that matter to us now are the cells to the left of 3^1. So here is the table.

		$3^{1/8}$	$3^{1/4}$	$3^{1/2}$	3^1	3^2	3^3	3^4	3^5	
		$1/9$	$1/3$	1	3	9	27	81	243	

"The cells on the right of 3^2 follow one pattern (the powers become one more each time). The cells on the left of 3^2 follow another pattern (the powers become half each time)."

"If the point is only filling the cells," I commented, "we have already succeeded. But then, we miss a very useful feature we had on the right of the table: to multiply two numbers on the top row (say, 3^{27} and 3^{10}) we can write the base and add the powers, so $3^{27} \cdot 3^{10} = 3^{37}$. This does not work on the left of the table. We might come up with a clever formula or method to find something like $3^{1/32} \cdot 3^{1/128}$ only based on knowing $1/32$ and $1/128$. But, whatever it might be, it is harder than just adding two numbers." (I was also aware of another, more serious cost: later we will want to use $3^{1/2}$ with another meaning, so that $3^{1/2} \cdot 3^{1/2} = 3$.)

"But I see other students want to say something. Perhaps they have seen another pattern that is not so broken."

The second pattern

"Look," replied the others, "when you move from 3^2 to 3^1, the power becomes one less. So, the next cell, should be 3^0."

				3^0	3^1	3^2	3^3	3^4	3^5	
		$1/9$	$1/3$	1	3	9	27	81	243	

"Nice!" I commented, "This holds also when we move from 3^3 to 3^2 or from 3^4 to 3^3 (see, 3 is one less than 4). Moreover, now we can add the powers in $3^0 \cdot 3^1$ to get to 3^1, that is what we want ($1 \cdot 3 = 3$; $3^0 \cdot 3^1 = 3^{0+1} = 3^1$). But then, what would be the power of the cell above $1/3$?"

"If we want to stick to the pattern," they answered, "the power of 3 in this case should be one less than zero. But we cannot have less than zero, can we?"

(And suddenly came a suggestion, that I am still pleasantly surprised about after 20 years.)

"Look!" yelled one of the students, "3^1 is just 3, 3^2 is two 3's multiplied together. We can represent these with $3^{\times 1}$ and $3^{\times 2}$. Then, as $^1/_3$ is a division, we can represent it with $3^{\div 1}$, and with $3^{\div 2}$ and so on!"

Since we were already using 3^2 for $3^{\times 2}$, we agreed to continue using the familiar notation on the right of the table.

$3^{\div 4}$	$3^{\div 3}$	$3^{\div 2}$	$3^{\div 1}$	3^0	3^1	3^2	3^3	3^4	3^5	
$^1/_{3^4}$	$^1/_{3^3}$	$^1/_{3^2}$	$^1/_3$	1	3	9	27	81	243	

What we did with the invented notation

This notation (that I accepted for the time being) had certain similarities with the standard notation that I had in mind, the most crucial one being that the rule for multiplying by adding the powers works the same on both sides of the table. For example, to find $3^{\div 27} \cdot 3^{10}$, we can write the base and *add* the powers: $3^{\div 17}$, that is $^1/_{3^{17}}$. This remains true regardless of the base. So, we started exploring the arithmetic of these objects with one general principle in mind: we wanted to keep the rules we knew for the powers intact. Multiplication of powers gave us addition (of our new objects). Division of powers gave us subtraction. The power of a power gave us multiplication. And division came as the inverse of multiplication.

$\frac{1}{3^{17}}$

The evaluation of the power of a power is worth explaining as it provides a mathematical reason for one of the most famous rules of negative numbers.

We want when finding the power of a power, to keep the base and multiply the powers. Let us apply this rule to $(a^{\div 4})^{\div 5}$. $(a^{\div 4})^{\div 5}$ *should be* $a^{(\div 4)\cdot(\div 5)}$. But, what is $(\div 4) \cdot (\div 5)$? Let's see.

$$(a^{\div 4})^{\div 5} = \frac{1}{(a^{\div 4})^5} = \frac{1}{a^{\div 20}} = a^{20}$$

So we *have to have* $(\div 4) \cdot (\div 5) = 20$.

We did all of these calculations in the class. At this stage, the main question for me (as the teacher) was why and how I should introduce the standard symbol, the minus sign.

Why –

Socially, shared signs are the key to successful communication. If each of us had idiosyncratic notations, soon no one would understand anyone else, and mathematics would become a decryption game. So, with every concept comes a set of agreed upon notations, one of them being -1 for denoting the solution of $s + 1 = 0$, and not, say, $\div 1$. But, is this a wise choice, considering that the sign denoting negative numbers will be the same as the sign of subtraction? Surprisingly, this ambiguous use of the sign – turns out to be the main reason for 'choosing' it to denote negative numbers.

Negative numbers are not usually encountered for the first time when exploring exponents. And historically, the *rules of signs* were first practiced in the algebraisation of arithmetic, before a general admission of negative numbers (Asghari, 2019). For example, once operations on natural numbers are understood, it becomes possible to combine them, and, for example, multiply binomials.

$$(a + b)(c + d) = ac + ad + bc + bd$$
$$(a + b)(c - d) = ac - ad + bc - bd$$
$$(a - b)(c - d) = ac - ad - bc + bd$$

None of these equalities needs any knowledge of negative numbers and each one can be justified geometrically. When multiplying $(a - b)(c - d)$, the term $+\,bd$ comes from the rule that 'minus times minus is plus' (*not* 'negative times negative is positive'). No knowledge of integers is required. It is all signs.

And we can apply the same rules to something like $(2 - 5)(3 - 7)$ that does not make sense in the realm of natural numbers, and for which the usual geometric models fail to represent. We can blindly apply the rules of signs to get a 'result':

$$(2 - 5)(3 - 7) = 2 \cdot 3 - 2 \cdot 7 - 5 \cdot 3 + 5 \cdot 7 = 6 - 14 - 15 + 35 = 12$$

Apart from the original multiplication, everything else in this chain has meaning in the realm of natural numbers. So, if $(2 - 5)(3 - 7)$ is going to have any meaning, any result, it should be equal to 12.

We could have performed the multiplication using my students' invented notation. $(2 - 5)(3 - 7) = (2 + \div 5)(3 + \div 7)$, because $(2 - 5)$ is the exponent of $a^2 \div a^5$, and so is $(2 + \div 5)$. The process of multiplication goes as follows:

$$(2 - 5)(3 - 7) =$$
$$(2 + \div 5)(3 + \div 7) =$$
$$2 \cdot 3 + 2 \cdot \div 7 + \div 5 \cdot 3 + \div 5 \cdot \div 7 =$$
$$6 + \div 14 + \div 15 + 35 =$$
$$6 - 14 - 15 + 35 =$$
$$12$$

With the sign –, we do not even need to write $(2 - 5)(3 - 7)$ as $(2 + -5)(3 + -7)$. We only need to know when we multiply the minus signs, we do not multiply them as subtraction signs; we multiply them as the signs preceding 5 and 7. Simply, the pre-integer knowledge of the signs suffices for all the calculations and there is no need to learn new rules.

Concluding the story

Thinking of negative numbers as minus numbers brings a procedural flexibility. However, pedagogically we should be cautious about thinking of negative numbers as minus numbers, as it might create conceptual obstacles when our students move to the algebra of variables (Asghari, 2019). Knowing why the sign – works and why it is procedurally better than the rival signs, may help students to see a negative number as a number in itself, not simply a whole number preceded by a minus sign. I later told my students to use $-a$ simply so they could communicate with the others mathematically. So, then I just said, "From now on, we use –, rather than ÷!" Now, with the benefit of hindsight, I could do a better job moving from our invented symbol to the standard one.

Note

[1] Online at http://www.corestandards.org/Math/Content/mathematics-glossary/glossary/

References

Asghari, A.H. (2019) Signed numbers and signed letters in algebra. *For the Learning of Mathematics* **39**(3), 13-16

Joseph, G.G. (2010) *The Crest of the Peacock: non-European Roots of Mathematics.* Princeton, NJ: Princeton University Press