

The “Rule of Signs” in Arithmetic

by Roger Howe and Solomon Friedberg

November 19, 2008

The “Rule of Signs” governing the multiplication of signed numbers, and especially the rule “A negative times a negative is positive,” has bewildered and distressed generations of students. In this note, we will justify these rules, and try to make them seem reasonable, perhaps even intuitive. We will begin with a theoretical treatment, and follow this with several justifications using concrete situations where the rules of signs are needed.

I. Theoretical Grounding of the Rules of Signs.

In the course of doing arithmetic, we gradually become aware of certain broad regularities that can be exploited in many situations to simplify computations. Indeed, use of these “rules” or “properties” is essentially unavoidable, since they have implicitly been built in to the standard algorithms for calculation, and indeed, they are exploited, again implicitly, in the very way we write numbers, our standard base ten (aka decimal) place value notation.

By gradual refinement over several centuries, and culminating in the late 19th century, mathematicians isolated a small collection of rules or properties that govern all of ordinary arithmetic, in the sense that all legitimate operations involving addition and subtraction, multiplication and division follow from these few rules. This collection is sometimes labeled the Field Axioms. Here we call them the *Rules of Arithmetic*. All the various “rules” commonly used for calculating can be deduced from the Rules of Arithmetic. Our goal here is to show how this is done for the “Rule of Signs.”

There are nine Rules: four govern addition, four govern multiplication, and one (the Distributive Rule) connects the two operations. (We have not said anything about subtraction or division. That is because, under these rules, subtraction is subsumed under addition, and division is subsumed under multiplication. We will say more about this after stating the Rules.) In stating them, it is assumed that we are dealing with some objects called “numbers.” It is assumed that we know how to combine any two numbers a and b by addition, to form their *sum*, which is another number, indicated by $a + b$. It is also assumed that we know how to combine any two numbers a and b by multiplication, to form their *product*, which is another number, indicated by $a \times b$.

1. Rules for Addition.

We begin with a discussion of the four basic rules for addition. These already are interesting and allow us to clarify issues that might cause trouble later on.

Ai) (Commutative Rule for Addition): For any two numbers a and b ,

$$a + b = b + a.$$

Aii) (Associative Rule for Addition): For any three numbers a , b and c ,

$$(a + b) + c = a + (b + c).$$

In this equation, the notation $(a + b) + c$ means that we form the sum of a and b , and then add this number to c . Similarly, the notation $a + (b + c)$ means that we form the sum $b + c$, and then add a to this number. The Associative Rule says that these two schemes for adding three numbers always give the same result.

Aiii) (Existence of Additive Identity (aka Zero)): There is a number 0 such that for any number a ,

$$0 + a = a.$$

Aiv) (Existence of Additive Inverse (aka “the negative” and “the opposite”)): For every number a , there is another number, called the *additive inverse* or *negative* of a , and denoted by $-a$, such that

$$(-a) + a = 0.$$

Here 0 is the zero element whose existence is asserted in Rule Aiii).

2. Consequences of the Addition Rules.

It can be instructive to see that several properties of addition that are familiar to us and that we expect to be true are in fact guaranteed by the four basic rules for addition. We can be confident that they are true without any extra assumptions. (This helps underscore the fundamental nature of the basic rules.) We will review some of the most important.

Uniqueness of zero: It is perhaps natural to expect that there could only be one possible zero element. A nice feature of these rules is that we do not need explicitly to assume the uniqueness of zero: it is a consequence of commutativity and the defining property of zero. For suppose that there are two candidates, 0 and $0'$, for the zero element; that is, both 0 and $0'$ satisfy the Existence of Zero (Rule Aiii). Then we have

$$0' = 0 + 0' = 0' + 0 = 0.$$

Here the first equation is Rule Aiii) for 0, the second equation is commutativity, and the third equation is Rule Aiii) for $0'$. Thus $0' = 0$. Note that here we are appealing to the general principle of *transitivity of equality*: *things both equal to a third thing are equal to each other*.

Cancellation: A useful consequence of the Existence of Additive Inverse (Rule Aiv)) is the *Cancellation Rule*:

For any numbers a , b and c , if

$$a + b = a + c, \quad \text{then} \quad b = c.$$

Indeed, if we are given that $a + b = a + c$, then we can add $-a$ to each side of this equation, to get $(-a) + (a + b) = (-a) + (a + c)$. (This basic fact, that *if two expressions are equal, then we may add the same thing to both sides and the results are equal*, is critical in deducing the consequences of the fundamental rules of arithmetic. It is a consequence of addition being well-defined, that is, there is only one answer for the sum of two given numbers.) Now we can simplify both sides using the rules. For example,

$$(-a) + (a + b) = ((-a) + a) + b = 0 + b = b.$$

Here the first equation uses Associativity (Rule Aii)), and the second equation follows from Existence of Additive Inverse (Rule Aiv)), and the third equation uses Existence of Zero (Rule Aiii)). We can show that $(-a) + (a + c) = c$ in similar fashion. Combining all these equations (and once again appealing to the general principle of transitivity of equality), we conclude that $b = c$.

In the argument for the Cancellation Rule, the key was to use the Existence of Inverse (Rule Aiv)), that provides us with additive inverses. However, when children begin to study arithmetic, they work in the whole numbers (aka, the non-negative integers), without the benefit of additive inverses. In that context, the Cancellation Rule can serve as a substitute for Rule Aiv).

Uniqueness of the additive inverse: Along with the uniqueness of zero, we probably expect that the negative of a number should be unique. More precisely, the negative, $-a$, should be the only number a' that can satisfy $a' + a = 0$, that is, that cancels out a when they are added. A third happy fact is that this also does not have to be assumed separately: it is another consequence of the four basic assumptions.

The uniqueness of $(-a)$ is easy to show using cancellation. If a' is any number such that $a' + a = 0$, then we have that

$$a + a' = a' + a = 0 = (-a) + a = a + (-a).$$

Cancelling the a from the first and last expressions, we get that $a' = (-a)$, as desired.

Zero is its own negative: Since everything has a unique additive inverse, it makes sense to talk about -0 , the additive inverse of zero. What is this? Using the Existence of Zero (Rule Aiii)) with $a = -0$ gives $0 + (-0) = -0$, while the Existence of Additive Inverse (Rule Aiv)) with $a = 0$ gives $(-0) + 0 = 0$. Using commutativity (Rule Ai)), we conclude that $-0 = 0$.

The inverse of the inverse: A fifth consequence of the four basic rules for addition is the reciprocal relationship between a number and its additive inverse: a number is the additive inverse of its additive

inverse; or a number is the negative of its negative. This is basically a consequence of commutativity (Rule Ai): if $(-a) + a = 0$, then also $a + (-a) = 0$ by commutativity. Thus, a satisfies the condition required of $-(-a)$ by the Existence of Additive Inverse (Rule Aiv)). Since we now know that Rule Aiv) defines the additive inverse uniquely, this implies that

$$-(-a) = a.$$

This equation may look related to the Rule of Signs, and as we will see, it is. However, there is no multiplication going on here. We are operating solely in the realm of addition, and this is a fact about an operation of addition satisfying rules Ai) through Aiv). It is a combination of the fact that addition is commutative (Rule Ai)), so that the relation between a number and its negative is symmetric, and the uniqueness of the additive inverse (which is a consequence of cancellation, which is a consequence of the Existence of Additive Inverse (Rule Aiv)!).

The additive inverse of a sum: If a, b are numbers, then the additive inverse of $a + b$ is equal to the sum of the additive inverse of a and the additive inverse of b , or in symbols, $-(a + b) = (-a) + (-b)$. The proof is similar in spirit to those above; the reader is encouraged to check this but we will not give the proof here.

Subtraction as addition of the additive inverse: When we have an addition governed by the four rules Ai) through Aiv), we no longer need subtraction.

Subtraction is created to solve equations like $a + x = c$. That is to say, $x = c - a$ is the solution to this equation¹. However, we can solve this equation using $(-a)$: if $a + x = c$, then $(-a) + (a + x) = (-a) + c$. But we can write $(-a) + c = c + (-a)$ by commutativity, and we can also deduce, as we have in several places above, that $(-a) + (a + x) = ((-a) + a) + x = 0 + x = x$. Thus, we can conclude that $x = c + (-a)$, which means that $c + (-a)$ is the number we usually call $c - a$. In other words, addition of the additive inverse of a does the same thing as what we are used to calling “subtraction of a .”

Once we understand subtraction as addition of the additive inverse, we see that we do not need separate rules for subtraction. For example, for all numbers a , we have $0 - a = 0 + (-a) = -a$ (the last since 0 is the additive identity), and for all numbers a, b we have $a - b = a + (-b) = (-(-a)) + (-b) = -((-a) + b) = -(b - a)$. These equalities follow from the consequences above.

In some sense, subtraction is a stopgap, an idea introduced to cope with the fact that when we (each of us as individual children, and historically, our ancestors) first met arithmetic, it was only about the whole numbers (or even only the strictly positive whole numbers - excluding zero). We did not have Rule Aiv): additive inverses do not exist within the system of whole numbers. Nevertheless, we needed a way to talk about doing what we can now describe as adding the additive inverse, and subtraction responds to this need. After years of practice, the idea of subtraction is pretty well ingrained in students’ minds, and the transition to the viewpoint that it is just an appendix to addition, while liberating, is not necessarily easy to accept. Instruction should take note of the difficulty of this essential transition.

3. Rules for Multiplication.

The rules for multiplication are essentially parallel to the rules for addition. However, it should be noted that the justification, and even the truth, of these rules, is much less immediate than for the related rules for addition. Indeed, even the interpretation of multiplication as an operation tends to be more mysterious than the interpretation for addition. The initial explanation, that “multiplication is repeated addition,” is inadequate for dealing with multiplication by numbers other than whole numbers. And even for whole numbers, it does not supply images that can help students think about multiplication and how it should behave. Instruction that discusses these rules should not treat them as self-evident and unworthy of discussion, but should use them to bolster the intuitive understanding of multiplication.

Although multiplication is a deeper operation than addition, our discussion of it and its rules will be briefer than for addition, since they are hardly needed for the Rule of Signs.

¹ Subtraction is first modeled with whole numbers by the concept of “taking away”; $7 - 4$ apples is the number of apples left if one starts with 7 apples and takes away 4. Of course if one then gives them back, one has 7 again, or $(7 - 4) + 4 = 7$. More generally, $(c - a) + a = c$. So $c - a$ is the solution to the equation $x + a = c$, or, by commutativity, to $a + x = c$.

Here are the rules.

Mi) (Commutative Rule for Multiplication): For any two numbers a and b , we have

$$a \times b = b \times a.$$

Mii) (Associative Rule for Multiplication) For any three numbers a , b and c ,

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

Miii) (Existence of Multiplicative Identity (aka One)): There is a number 1 such that for any number a ,

$$1 \times a = a.$$

Miv) (Existence of Multiplicative Inverse (aka “the reciprocal”)): For every non-zero number a , there is a another number, called the *multiplicative inverse* or *reciprocal* of a , and written as $\frac{1}{a}$ (or sometimes also as a^{-1}), such that

$$\frac{1}{a} \times a = 1.$$

4. Consequences of the Multiplication Rules.

Since the rules for multiplication are almost completely analogous to the rules for addition, the consequences that we noted for addition hold likewise for multiplication: the multiplicative identity (i.e., 1) is unique, there is a cancellation rule (if $a \times b = a \times c$ and a is non-zero, then $b = c$), the reciprocal (multiplicative inverse) is unique, the reciprocal of the multiplicative identity is itself, the reciprocal of the reciprocal is the original number, the reciprocal of a product is the product of the reciprocals, and division is subsumed in multiplication, in the sense that division by a is the same as multiplication by the reciprocal of a . In saying all these things, it is important to require that a be non-zero, because of the caveat in Rule Miv). This is the lone exception to the broad parallelism between the two sets of rules. It is an absolutely essential difference, as we will shortly see.

5. Connecting Multiplication and Addition: the Distributive Rule.

The above two sets of four rules capture the essential features of the two operations separately. It might be expected that to combine them into the full system of arithmetic would require a large raft of further rules. Remarkably, only one more is needed. Apparently, a system of numbers that supports two separate operations is already a fairly rigid object, and only the minimal coordination between the two operations, as provided by the Distributive Rule, is enough to capture the key aspects of arithmetic.

D) (Distributive Rule): For any three numbers a , b and c ,

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

To make this statement completely clear, we should probably place the expressions $a \times b$ and $a \times c$ in parentheses, to indicate that on the right hand side, we are supposed to compute the products $a \times b$ and $a \times c$, and then add them together. However, we are relying on the standard order-of-operations conventions, in which one always performs any indicated multiplications before adding, unless a different order is specified by means of parentheses. This is the case on the left side of the identity, where the parentheses around $b + c$ indicate that we are supposed to perform this addition first, and then multiply the result by a .

Note that the commutativity of multiplication (Rule Mi)) combined with the Distributive Rule allows us also to assert that $(b + c) \times a = b \times a + c \times a$ – we have distributivity on the left as well as on the right. If multiplication were not commutative, one would have to assume distributivity on the right (which is the more precise name for Rule D)) and distributivity on the left, separately.

We do not need to assume a separate distributive law for multiplication over subtraction; rather such a statement is a consequence of this law and the link between subtraction and additive inverses. We will derive it later.

6. Some Consequences of the Distributive Rule.

An immediate consequence of the the Distributive Rule is that the Cancellation Rule fails very badly for multiplication by zero.

Zero times anything is zero: For any number a , we have

$$0 \times a = 0.$$

Indeed, by the defining property of 0, Rule Aiii), we know that for any number b , $0 + b = b$. If we multiply both sides of this by a (on the right), we find that $(0 + b) \times a = b \times a$. The Distributive Rule lets us expand the left hand side of this equation, producing $0 \times a + b \times a = b \times a = 0 + b \times a$. The second equality here again uses the defining property of 0. Now using (additive) cancellation of $b \times a$, we conclude that $0 \times a = 0$. (We have cancelled $b \times a$ on the right, whereas our statement of the Cancellation Rule was for cancellation on the left. Of course, they are both allowed, since addition is commutative. A completely methodical discussion of arithmetic would have noted cancellation on the right explicitly.)

This property of zero in multiplication shows that the cancellation rule for multiplication by zero fails completely: $0 \times a = 0 \times b$ does *not* imply that $a = b$, rather it is true for any a, b , whether equal or not. We recall that the existence of a multiplicative inverse of a number c , which is equivalent to being able to divide by c , implies a cancellation rule for multiplication by c . (See the analogous discussion for the cancellation rule in addition.) That is to say, dividing by a number is only possible when there is a cancellation rule for multiplication by that number. Thus, the utter failure of cancellation for multiplication by zero, which is already visible in whole number arithmetic, shows that we could not possibly introduce a reciprocal of zero, or in other words, there is no hope of dividing by zero.

A second consequence is a link between additive inverses and multiplication: we can obtain additive inverses by multiplication by -1 .

Minus one times anything is its additive inverse: For any number a , we have

$$(-1) \times a = -a.$$

Indeed, $((-1) + 1) \times a = 0 \times a = 0$. In the last equality we have used the result in the previous paragraph that zero times anything is zero. But, using the Distributive Rule, Rule D), to expand the first expression, this gives $(-1) \times a + 1 \times a = 0$. Hence, by Existence of Multiplicative Identity (Rule Miii)), $(-1) \times a + a = 0$. Comparing with the Existence of Additive Inverse (Rule Aiv)), and recalling that the additive inverse is unique, we deduce that $(-1) \times a$ is the additive inverse of a , that is, $(-1) \times a = -a$.

7. The Rule of Signs.

The third consequence of the Distributive Rule that we will discuss is what is usually taken as the Rule of Signs.

For any numbers a and b , we have

$$a \times (-b) = (-a) \times b = -(a \times b); \quad \text{and} \quad (-a) \times (-b) = a \times b.$$

To show these identities are true, start with $(-b) + b = 0$. Multiplying by a gives $a \times ((-b) + b) = a \times 0 = 0$ since zero times anything is zero. Expanding the first expression using the Distributive Rule lets us conclude that $a \times (-b) + a \times b = 0$. This says that $a \times (-b)$ is the additive inverse of $a \times b$. We know that the additive inverse is unique. (This is why we wanted to be sure of that fact!) This means that $a \times (-b) = -(a \times b)$. The equality $(-a) \times b = -(a \times b)$ then follows by switching the roles of a and b , and using the commutativity of multiplication.

Finally, the killer identity $(-a) \times (-b) = a \times b$ follows from these relations and from the fact that the negative of the negative is the original number. We have

$$(-a) \times (-b) = -(a \times (-b)) = -(-(a \times b)) = a \times b.$$

The first equality here follows from the equation $(-a) \times b = -(a \times b)$ by substituting $-b$ for b , and the last step uses “the negative of the negative is the original.”

We also remark that as a special case, $(-1) \times (-1) = 1$. In fact, since $(-1) \times a = -a$, all the rules of signs can be deduced from this and the equations $1 \times (-1) = -1$, $(-1) \times 1 = -1$, $(-1) \times (-1) = 1$ via the associative and commutative properties of multiplication.

OK, maybe this is not the demonstration to show students when they are first learning about negative numbers. However, it does make the point that basic facts about addition as an operation, independent of any discussion of multiplication, together with the Distributive Rule, give us no choice about how to multiply negative numbers. We should also take note that this does not depend on the existence of a multiplicative identity or the existence of a multiplicative inverse, or even on the associativity of multiplication. We did use the commutativity of multiplication, but this too could be avoided. The key ingredients are the rules for addition, and the Distributive Rule.

Here at the end, we should also confess that this formulation of the Rule of Signs really has nothing to do with positive and negative numbers. It simply says that the product of a number with the additive inverse of another number is the additive inverse of the product of the two original numbers. Repeating this twice, once on each number, says that the product of the additive inverses of two numbers is the same as the product of the two original numbers. It doesn't matter whether the original numbers are positive or negative, or even if “positive” and “negative” mean anything. (See examples 4 and 5 in Section II below.) Nevertheless, when a and b are positive integers, then $-a$ and $-b$ are negative integers, and our formula does say that their product is the same as ab , which is positive.

The distributive law for differences: We give one consequence of the Rule of Signs, namely the distributive law for multiplication over subtraction:

For any numbers a , b and c , we have

$$a \times (b - c) = a \times b - a \times c; \quad \text{and} \quad (a - b) \times c = a \times c - b \times c.$$

To prove this, we make use of the link between subtraction and additive inverses along with the the standard Distributive Law, Rule D) above. We have $a \times (b - c) = a \times (b + (-c)) = a \times b + a \times (-c) = a \times b + (-(a \times c)) = a \times b - a \times c$. We have used the first part of the Rule of Signs here. The second distributive law above is proved in a similar way by using the left distributive law, or by using commutativity.

In conclusion, we again emphasize the theme that *a small number of rules, together with mathematical reasoning, determines the other properties of arithmetic.*

II. The Rule of Signs in Contexts.

The discussion in part I was highly theoretical. However, there is a wide variety of concrete situations which require multiplying signed numbers and which can be invoked to justify the Rule of Signs. Here is a selection.

1. Bank Balances.

You can keep track of your bank balance as you receive credits (for example, from checks you deposit in your account) and as you make withdrawals (for example, by using your debit card). If you receive a credit, your bank balance goes up. This is a positive change. If you make a withdrawal your bank balance goes down. This is a negative change.

If you receive a credit of \$5, your bank balance increases or changes by \$5. If you make a withdrawal of \$5, your bank balance goes down by \$5, or changes by $-\$5$.

If you receive three credits for \$5 each, your balance changes by $3 \times \$5 = \15 . If you make three withdrawals of \$5 each, your bank balance decreases by $3 \times \$5 = \15 . In other words, it changes by $3 \times (-\$5) = -\15 .

If three credits to your account of \$5 each are cancelled (for example, because the checks you deposited did not clear), then your actual bank account is \$15 less than expected. This is a change of $(-3) \times \$5$ from what you expected, so $(-3) \times \$5 = -\15 .

Finally, if you made three withdrawals of \$5 each and your withdrawals are cancelled (for example, the items you purchased were defective and the merchant cancelled the debits) then you have \$15 more than you expected to when you authorized the withdrawals, so $(-3) \times (-\$5) = \15 .

Although it may not seem that we used the Distributive Rule here, it is implicit in the determination of the correct increments in the bank balance. When we attributed a \$15 increment from cancelling three withdrawals, we were imagining that the three withdrawals had been made, resulting in a \$15 decrease, so that cancelling them, which resulted in a net change of \$0, could be interpreted as a \$15 increase from the situation that would have resulted from making the withdrawals. Thus, our argument was a veiled use of the relations $0 = 0 \times (-5) = (3 + (-3)) \times (-5) = 3 \times (-5) + (-3) \times (-5)$. This kind of argument is also hiding behind each scenario in the bank account example above, and the weight gain and loss example below.

2. Weight gain and loss.

We can imagine a person gaining or losing weight, depending on what they eat.

Suppose for some time the person has been eating a lot, and is gaining 5 pounds per week. Then in 3 weeks, the person will have gained fifteen pounds: $3 \times 5 = 15$.

On the other hand, three weeks ago, the person weighed fifteen pounds less: $(-3) \times 5 = -15$.

Suppose that our person switches to a successful diet, and begins losing 5 pounds per week. Then in three weeks the person will lose 15 pounds: $3 \times (-5) = -15$.

Later, when well into the diet, the person can see that 3 weeks ago, their weight was 15 pounds more than it is now: $(-3) \times (-5) = 15$.

Again, we point out the implicit use of distributivity. For example, in justifying the equation $(-3) \times 5 = -15$, we were actually simply computing that the person should have been 15 pounds lighter 3 weeks ago in order to arrive at their current weight now by gaining 5 pounds per week. In other words, our essential claim was that $(-3) \times 5 + 3 \times 5 = 0$, which is justified formally by the Distributive Rule, in the manner we have seen.

3. Area.

Multiplication can be interpreted in terms of area, and several of our rules are easy to visualize when we do so. The product $a \times b$ of two positive numbers a and b is the area of a rectangle of height a and base b . See Figure 1. If we turn the rectangle on its side, it has the same area, but this rotated rectangle has height b and base a . Thus we obtain the Commutative Rule (Mi): $a \times b = b \times a$.

It is also easy to interpret the Distributive Rule in terms of area, and doing so is valuable in seeing that this rule has a clear and concrete physical meaning. If the height of a rectangle is a , but the base is $b + c$, then the rectangle can be cut by a vertical line into two smaller rectangles, both of height a , and one of base b and the other of base c . See Figure 2. This illustrates that $a \times (b + c) = a \times b + a \times c$.

If we pursue this line of thinking, it is easy to imagine dividing a rectangle by several lines, some vertical and some horizontal, and obtaining extended versions of the Distributive Rule. For example, suppose we have a rectangle divided by a horizontal line into two subrectangles of heights a and b (so the total height of the rectangle is $a + b$). Suppose further that the rectangle is divided by a vertical line into two subrectangles of bases c and d (so the total length of the base of the rectangle is $c + d$). These two lines together partition the rectangle into four subrectangles of size $a \times c$, $a \times d$, $b \times c$ and $b \times d$. This provides us with a slightly extended version of the Distributive Rule, a recipe for multiplying two binomials:

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

See Figure 3.

It is easy to see how this works when we are multiplying positive quantities. However, we can also adapt the area interpretation to cope with signed numbers, at least to some extent. For example, think of $3 \times (8 - 2)$. We can think of a 3 by 8 rectangle, with area 24, and then we can imagine cutting off a 3 by 2 rectangle from one side, leaving a 3 by 6 rectangle. This illustrates $3 \times 6 = 3 \times (8 - 2) = 3 \times (8 + (-2)) = 3 \times 8 + 3 \times (-2) = 3 \times 8 - (3 \times 2)$, leading to the interpretation that $3 \times (-2) = -(3 \times 2) = -6$. See Figure 4.

We can do the same kind of thing with the height. If $3 = 7 - 4$, then we can think of $(7 - 4) \times 6$ as gotten by taking a rectangle of height 7 and base 6, and cutting off the rectangle of height 4 at the top. This will leave a rectangle of height 3 at the bottom, illustrating $(7 - 4) \times 6 = (7 + (-4)) \times 6 = 7 \times 6 + (-4) \times 6 = 7 \times 6 - (4 \times 6) = 3 \times 6$. This leads to the conclusion that $(-4) \times 6 = -(4 \times 6) = -24$. See Figure 5.

We can try to combine both of these changes at once. Writing $6 = 8 - 2$ and $3 = 7 - 4$, the Distributive Rule would tell us that

$$(7 - 4) \times (8 - 2) = (7 + (-4)) \times (8 + (-2)) = 7 \times 8 + 7 \times (-2) + (-4) \times 8 + (-4) \times (-2).$$

We have seen that $7 \times (-2) = -(7 \times 2)$, and that this term corresponds to cutting off a 7 by 2 rectangle from the side of the large 7 by 8 rectangle. Likewise, we have seen that $(-4) \times 8 = -(4 \times 8)$, and that this term corresponds to cutting off a 4 by 8 rectangle from the top of the 7 by 8 rectangle. Now we have cut off both a 7 by 2 rectangle from the side, and 4 by 8 rectangle from the top, leaving only a 3 by 6 rectangle of area 18. However, $56 - 14 - 32 = 10$. We appear to be missing 8 units of area. Reflection reveals that, in cutting off the two smaller rectangles from the large rectangle, we have left the lower left 3 by 6 rectangle, but also, we have cut off the upper right 4 by 2 rectangle twice, since this upper right rectangle is part of both rectangles cut off. Therefore, to get the right value for the area of the lower left rectangle, we must add back in the area of the upper right rectangle. This is the function of the term $(-4) \times (-2)$, which therefore should contribute positively to the sum: $(-4) \times (-2) = 8 = 4 \times 2$. With this interpretation, the equation resulting from the Distributive Rule for this situation is

$$18 = 3 \times 6 = 7 \times 8 - 7 \times 2 - 4 \times 8 + 4 \times 2 = 56 - 14 - 32 + 8,$$

which is valid. See Figure 6.

Thus, the geometric interpretation of the Distributive Rule with signed numbers also leads us to the conclusion that $(-a) \times (-b) = a \times b$ for positive numbers a and b .

4. Clock arithmetic.

The rule of signs is valid even in systems of arithmetic different from our usual one. Consider for example the numbers on a standard clock face. If one starts at 7 o'clock and passes 7 hours, the time is not 14 o'clock but rather 2 o'clock. To distinguish this system of arithmetic from our usual one, we write clock numbers with an overscore. So for clock arithmetic, we have $\overline{7} + \overline{7} = \overline{2}$. Since going forward 12 hours results in the same time, the number 12 functions as the additive identity: $\overline{12} = 0$. The rules of addition hold. For example, since $\overline{1} + \overline{11} = \overline{12} = 0$, the additive inverse of $\overline{1}$ is $\overline{11}$: $-\overline{1} = \overline{11}$. We can similarly define the product of two numbers \overline{a} and \overline{b} to be the clock time obtained by starting at 12 o'clock and going forward $a \times b$ hours. For example, $\overline{7} \times \overline{5} = \overline{11}$ since 35 hours after 12 o'clock gives 11 o'clock ($35 = 2 \times 12 + 11$), and $\overline{11} \times \overline{11} = \overline{1}$ since $11 \times 11 = 121 = 10 \times 12 + 1$; going around a clock 121 hours starting at 12 o'clock, one ends up at 1 o'clock. One can see that (Mi), (Mii) and (Miii) hold (and the multiplicative identity is just $\overline{1}$), but in fact (Miv) does not; not every number has an inverse². Also D) holds. Since we did not need (Miv) to prove the law of signs, in fact it holds for this system of arithmetic. For example, the key identity $(-\overline{1}) \times (-\overline{1}) = \overline{1}$ is true here since $(-\overline{1}) \times (-\overline{1}) = \overline{11} \times \overline{11} = \overline{1}$. And in general, $(-a) \times (-b) = a \times b$.

5. Complex numbers.

The Rule of Signs continues to hold even when there is no notion of positive or negative. One example is the clock arithmetic example immediately above. Another, important, example comes from Complex Numbers. The complex number i , which satisfies $i \times i = -1$, is neither positive nor negative. (Indeed, we've just seen that the product of two positives and the product of two negatives are both positive.) However, the complex numbers do satisfy the Rules of Arithmetic, and so for any two complex numbers a, b , we have $(-a) \times (-b) = a \times b$. For example, $(-i) \times (-i) = i \times i = -1$. So the complex numbers i and $-i$ are both square roots of -1 .

6. Extending the multiplication table.

Consider the multiplication table for non-negative integers. By this we mean the infinite array of numbers in the positive quarter plane, located at the integer points (m, n) , with the number placed at (m, n) being mn , the product of m and n . See Figure 7. Extending multiplication to all integers would result in assigning numbers to all integer points in the whole plane, not just the positive quarter plane.

² For example, the only multiples of $\overline{6}$ are $\overline{6}$ and $\overline{12}$. Since one can not get $\overline{1}$ as a multiple of $\overline{6}$, the number $\overline{6}$ does not have a multiplicative inverse.

What is a reasonable way to do this? One can look for patterns in the multiplication table for positive integers, and extend the table so that these patterns are preserved. One prominent pattern is this: if you move a single step to the right in one horizontal row, the product increases by the same amount, irrespective of where you are in the row. More precisely, if it is the row R_m , consisting of points (n, m) with fixed second coordinate equal to m , then with each step to the right, the product increases by m . (Note that this amounts to the identity $(n + 1) \times m = n \times m + m$. This is a special case of the Distributive Rule. However, one can show by mathematical induction that this special case implies the full rule, at least for non-negative integers. Also, one can show that this rule in fact completely determines the multiplication table.)

We are interested in moving to the left, from the positive quadrant to the upper left quadrant where the first coordinate is negative. Thus, we rephrase the pattern above by saying, that if you move one step to the left in the row R_m , then the product decreases by m . This is a very simple pattern which we can repeat again and again as we move to the left. Doing so, we find successively that we should define:

$$(-1) \times m = 0 \times m - m = -m,$$

$$(-2) \times m = (-1) \times m - m = -m - m = -(m + m) = -(2 \times m),$$

$$(-3) \times m = (-2) \times m - m = -(2 \times m) - m = -(2 \times m + m) = -(3 \times m),$$

and in general, we find that

$$(-n) \times m = -(n \times m)$$

for any non-negative integers n and m . See Figure 7.

If we repeat this reasoning process in thinking about moving up and down a fixed column C_n consisting of points with fixed first coordinate equal to n , we can extend the multiplication table to the quadrant where the first coordinate is positive and the second coordinate is negative. We will arrive at the analogous result:

$$n \times (-m) = -(n \times m)$$

for any non-negative integers m and n . See Figure 8.

After completing both the extension procedures sketched above, we will have a table defining the product $n \times m$ in the three quadrants where at least one of n or m is positive. This definition satisfies the sign rule $(-n) \times m = -(n \times m) = n \times (-m)$ for all positive n and m . See Figure 9.

We can use the same idea to extend the definition of multiplication to the lower left quadrant where both coordinates are negative. Actually, there are now two ways to do this. We can extend down from the upper left quadrant by moving down columns. We can check that in a column C_{-n} (where $n > 0$), each time we move down one step, the entry increases by n . If we continue this rule down, into the lower left quadrant, we obtain a candidate for an extension of multiplication to all integers. We have $(-n) \times 1 = -n$, then $(-n) \times 0 = -n + n = 0$, then $(-n) \times (-1) = 0 + n = n$, then $(-n) \times (-2) = n + n = n \times 2$. Continuing like this, we will find that $(-n) \times (-m) = n \times m$ for any non-negative integers n and m . See Figure 10.

We can also extend to the lower left quadrant by moving to the left along rows from the lower right quadrant. A reasoning process like that of the previous paragraph shows that we again get the result that $(-n) \times (-m) = n \times m$. That is, both extension methods agree, and both produce the standard Rule of Signs, and in particular, both say that a negative times a negative is positive.

One can extend multiplication from the natural numbers to the integers simply by making these definitions, which are motivated in that one is continuing the patterns we observed above. However, when one does so it is not apparent that rules such as the Associative Law for Multiplication and the Distributive Law continue to hold. They do, of course, but this needs to be checked.

7. Multiplication as stretching the number line.

Another geometric model for multiplication is provided by considering the transformations of the number line provided by multiplication by a fixed number a . Given a positive number a , we can define the transformation *dilation by a* on the number line by

$$D_a(b) = a \times b.$$

Here b is an arbitrary real number. See Figure 11 for an illustration of D_2 . In making this definition, we assume that we know how to multiply any number by a positive number. In particular, we assume that if $a > 0$ and $b > 0$, then $a \times (-b) = -(a \times b)$. This is the least controversial of the rules for multiplying positive and negative numbers. We are taking the point of view that this rule (“multiplying a negative by a positive is a negative”) is easy to accept, and that the rules that cause angst are the rules for multiplying any number by a negative, and especially, the rule for the product of two negative numbers.

If x and y are any real numbers with $x < y$, then $[x, y]$ denotes the interval of all points between x and y . It has length $y - x$. If we look at what D_a does geometrically, we can see that it stretches any interval by the factor a . The interval $[0, 1]$ goes to the interval $[0, a]$, which has length $a = a \times 1$. The interval $[1, 3]$ goes to the interval $[a, 3a]$, which has length $3a - a = 2a = a \times (3 - 1)$. This argument is completely general. The image of the interval $[x, y]$ under D_a is $[ax, ay]$, which has length $ay - ax = a(y - x)$ (by the Distributive Rule).³ Thus, the transformation D_a will stretch any interval by a factor of a . This is the rationale for the name “dilation by a .”

Can we extend this definition to include “dilation by a negative number”? To consider how we might do this, we note some salient properties of D_a for $a > 0$:

- i) D_a preserves the origin: $D_a(0) = 0$.
- ii) D_a takes 1 to a : $D_a(1) = a$.
- iii) D_a stretches the length of every interval by a factor of a .

These properties can easily be shown to determine D_a uniquely.

If a is positive, it turns out that we can define D_{-a} in a way that preserves the essence of these properties. It is simple: we set

$$D_{-a}(x) = -D_a(x).$$

What this means geometrically is that we dilate x by a , and then we reflect it in the origin. See Figure 12 for an illustration of D_{-2} . Since $x \rightarrow -x$ preserves distances between points on the real line, we can see that D_{-a} will have the following properties:

- i) D_{-a} preserves the origin.
- ii) D_{-a} takes 1 to $-a$: $D_{-a}(1) = -a$.
- iii) D_{-a} stretches the length of every interval by a factor of $a = |-a|$.

Here $|a|$ denotes the absolute value of a .

These properties together, and their parallelism with the properties for multiplication by a when $a > 0$, make it eminently plausible that D_{-a} should be the transformation resulting from multiplication by $-a$. That is, we are led to propose the definition that

$$(-a) \times x = D_{-a}(x).$$

If we do accept this, then we see we must also accept that a negative times a negative is positive, since according to this definition, D_{-a} flips the number line over the origin, thereby converting negative numbers to positive ones. Algebraically, this equation yields that for a, b positive, $(-a) \times (-b) = D_{-a}(-b) = -D_a(-b) = -(a \times (-b)) = -(-(a \times b)) = a \times b$.

We can use dilations to view multiplication another way. The Associative Rule for Multiplication states that $a(bx) = (ab)x$ for all numbers a, b, x , and it follows that

$$D_a D_b = D_{ab}.$$

That is, the product of two numbers a and b is obtained by composing their dilations. From this point of view the identity $(-1) \times (-1) = 1$, from which the Rule of Signs follows, has the satisfying interpretation that carrying out two successive flips of the line across zero amounts to doing nothing at all - the identity. (The

³ If addition of real numbers has been interpreted geometrically using the number line as the process of “arrow addition” (aka “vector addition” and “concatenation of directed line segments”), then this interpretation can be used to verify the Distributive Rule.

rule ‘negative times negative is positive’ states that the result of two successive dilating flips is a dilation by some positive number, with no flip.)

As a concluding remark, we note that this geometric picture of multiplication of real numbers in terms of dilations of the number line is compatible with the geometric interpretation of multiplication of complex numbers. When complex numbers are pictured as points in the plane, multiplication by a complex number becomes a stretching by the absolute value of the number, combined with a rotation by the argument of the number. In this scheme, negative real numbers appear on the negative x -axis, and have an argument of 180° : they produce rotations of 180° of the whole plane. These rotations preserve the real axis, and there act as reflections in the origin, just as defined above. Thus the Rule of Signs can be seen as a first instance of the geometry of multiplication of complex numbers. The principles we teach in the early and middle grades are the foundation on which a vast edifice rests.