

Why we cannot divide by zero

Michael J. Neely

University of Southern California

<http://www-bcf.usc.edu/~mjneely>

These notes discuss why we cannot divide by 0. The short answer is that 0 has no multiplicative inverse, and any attempt to define a real number as the multiplicative inverse of 0 would result in the contradiction $0 = 1$. Some people find these points to be confusing. These notes may be useful for anyone with questions about dividing by 0. In order to ask *why* the division operation $1/0$ is not defined, it is important to understand where division comes from in the first place. Thus, the first part of these notes discuss axioms of arithmetic. Readers looking for quick insight and intuition, without all the mathematical overhead, can skip to the brief sections *D*, *E*, *F*.

A. Axioms

Arithmetic starts by assuming there are objects called *real numbers*. It assumes these real numbers can be manipulated by operations of *addition* and *multiplication* that take two real numbers and produce a third:

- Addition: For any two real numbers a and b , there is a real number $a + b$.
- Multiplication: For any two real numbers a and b , there is a real number ab .

These addition and multiplication operations are assumed to satisfy certain axioms, called *axioms for a field*. The axioms include standard commutative, associative, and distributive laws for arithmetic. The complete list of axioms is below:¹

- 1) Commutative law: $a + b = b + a$ and $ab = ba$.
- 2) Associative law: $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$.
- 3) Distributive law: $a(b + c) = ab + ac$.
- 4) Existence of 0: There is a number “0,” called the *additive identity*, that satisfies $a + 0 = a$ for all real numbers a .
- 5) Existence of 1: There is a number “1,” called the *multiplicative identity*, that satisfies $a1 = a$ for all real numbers a .
- 6) Additive inverse: For every real number a , there is a real number $-a$, called the *additive inverse of a* , that satisfies $a + -a = 0$.
- 7) Multiplicative inverse: For every nonzero real number a , there is a real number a^{-1} , called the *multiplicative inverse of a* , that satisfies $aa^{-1} = 1$.
- 8) $0 \neq 1$.

The axioms cannot be proven: They are a short list of properties that we intuitively expect numbers to satisfy. However, the axioms are *useful* in that, from them, important additional facts about numbers can be proven. An important theorem that can be proven directly from the axioms is that multiplying any real number by 0 produces 0 as the result.

Theorem 1: For every real number x , the following equation is true:

$$0x = 0$$

Proof: See Appendix A for the standard proof. □

B. Criticizing the multiplicative inverse axiom

One immediately notices an asymmetry in the way the additive and multiplicative inverse axioms above are stated (axioms 6 and 7 in the above list). The additive inverse axiom holds for *all real numbers*, while the multiplicative inverse axiom holds only for *all nonzero real numbers*. Why is 0 excluded? This is not fair! For a direct answer to this question, the reader can skip to Section *D*. Readers who prefer the question “*why can’t we divide by zero?*” must admit that this question cannot even be asked until we define division, which is done in the next section.

C. Subtraction and division defined as inverse operations

The axioms are stated in terms of addition and multiplication operations. Operations of subtraction and division are defined *after* the axioms are stated. Here are the standard definitions.

- Subtraction: For two real numbers a , b , the subtraction $a - b$ is defined as $a + -b$.
- Division: For two real numbers a , b , with b nonzero, the division a/b is defined as ab^{-1} .

In particular, $1/2$ is defined as 2^{-1} and $1/5$ is defined as 5^{-1} . Thus, $1/0$ is not defined, simply because there is no 0^{-1} . That is, 0 is not defined to have a multiplicative inverse.

¹The 8th axiom “ $0 \neq 1$ ” can equivalently be replaced by the assumption “there exists a nonzero real number.” Neither of these statements can be proven without the other, since the case of all numbers being 0 is consistent with the 7 other axioms. Strictly speaking, the real numbers require more structural assumptions to distinguish them from other *fields* that satisfy the 8 axioms. That additional structure is not relevant to these notes and hence is omitted.

As an example, if $bd \neq 0$ then the expression $\frac{ad+bc}{ad}$ is defined as:

$$\frac{ad+bc}{bd} = (ad+bc)(bd)^{-1}$$

This is the multiplication of the “numerator” $(ad+bc)$ and the multiplicative inverse of the “denominator” bd . The value $(bd)^{-1}$ is well defined since bd is assumed to be nonzero. It can be proven from the axioms that $bd \neq 0$ if and only if $b \neq 0$ and $d \neq 0$. Thus, if $bd \neq 0$ then both b^{-1} and d^{-1} exist. This can be used to derive standard rules for manipulating fractions such as (see Appendix B for the derivation):

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

which holds if $b \neq 0$ and $d \neq 0$ (else, one or more of the expressions are not defined). Another standard rule is:

$$\frac{1}{b/c} = \frac{c}{b}$$

which holds if $b \neq 0$ and $c \neq 0$ (else, one or more of the expressions are not defined).

D. The contradiction if 0^{-1} is defined

From the definition of division, the following questions are equivalent:

- Why can't we divide by 0?
- Why is the expression a/b undefined if $b = 0$?
- Why is the expression $1/b$ undefined if $b = 0$?
- Why doesn't 0 have a multiplicative inverse?

Let's focus on the last question. Why don't we define a particular number to be the multiplicative inverse of 0? Let's suppose we choose a particular real number z that we define as the multiplicative inverse of 0, so that $z = 0^{-1}$. Thus, by definition of a multiplicative inverse:

$$0z = 1$$

However, we also know from Theorem 1 that $0z = 0$. This means $1 = 0$, a contradiction. Thus, we cannot assign any real number to be 0^{-1} without reaching a contradiction in our pre-established facts.

E. Let's make z a non-real number

We could define $z = 0^{-1}$ to be a new symbol, not a real number. But in that case, we have to admit that this new symbol z may not obey the rules of arithmetic that real numbers do.

Once we admit that z may not obey the rules of arithmetic, we have the complicated task of figuring out *which* rules it does satisfy without leading to a contradiction. Even if we did that, we would have to figure out a reason *why* we are introducing such a number z . What is the use of it?

Since there is no apparent usefulness of such a non-real number z , it is easier to simply avoid trying to define it. Then, we just admit that 0 has no multiplicative inverse.

F. Intuition on reversibility

Here is an intuitive reason *why* 0 should not have a multiplicative inverse: An invertible operation should be “reversible” in the following sense:

If I take a number x and multiply by 5, I get $5x$. I can “reverse this” by multiplying by 5^{-1} (equivalently, dividing by 5) to get x back. Multiplying by 5 does not lose any information about the original number x . For example, if x is unknown, but I tell you I multiply x by 5 and the result is 10, you can determine what x was just by solving the equation $5x = 10$. The invertible operations of multiplying and dividing by 5 are like a photocopy machine that shrinks or expands a document by a factor of 5 depending on which button is pushed. We can shrink and expand as many times as we like, with no loss of information.

Now let's take an unknown number x and multiply by 0. The result is 0. I cannot recover x , all I have is 0. The equation $0x = 0$ does not offer any information about x . The photocopy machine reduced my document to nothing! *Multiplying by 0 is a noninvertible operation!* We cannot go back by “dividing by 0.”

G. Intuitive limit arguments

It can be shown that if n is a nonzero number, then $(1/n)^{-1} = (n^{-1})^{-1} = n$. Now let n be a very large positive number. We intuitively understand that $1/n$ approaches 0 as n is increased to ∞ . Thus, we intuitively expect $(1/n)^{-1}$ to approach 0^{-1} as n is increased to ∞ (assuming 0^{-1} is defined in some form). But $(1/n)^{-1} = n$, which has a limit of ∞ as n is increased to ∞ . This suggests that $0^{-1} = \infty$.

If this argument convinces you to define 0^{-1} as the non-real number ∞ , think again. By a similar argument it holds that $1/n$ approaches 0 as n is *decreased* to $-\infty$ (such as when n is chosen as -10000 and then -1000000000). Thus, intuitively one expects $(1/n)^{-1}$ to approach 0^{-1} as n is decreased to $-\infty$. But $(1/n)^{-1} = n$ has a limit of $-\infty$ as n approaches $-\infty$. This suggests that $0^{-1} = -\infty$. How can 0^{-1} be both ∞ and $-\infty$?

APPENDIX A—PROVING THEOREM 1

The axioms are stated with an implicit understanding of the notion of an *equation* such as $a(b+c) = ab+ac$. The equation means that the real number $a(b+c)$ is the same as the real number $ab+ac$. If an equation is “true,” then another true equation can be obtained by adding the same thing to both sides, or by multiplying the same thing to both sides.² This is used in the following proof. The proof also uses notation $x1$ to represent the multiplication of a real number x and the number 1.

Proof: (Theorem 1) We know from axiom 4 that:

$$1 + 0 = 1$$

Multiplying both sides by x gives:

$$x(1 + 0) = x1$$

Using the distributive law gives:

$$x1 + x0 = x1$$

Using the fact that $x1 = x$ gives:

$$x + x0 = x$$

Adding $-x$ to both sides gives:

$$-x + (x + x0) = -x + x$$

Using the associative law:

$$(-x + x) + x0 = -x + x$$

Using the fact that $-x + x = 0$ gives:

$$0 + x0 = 0$$

Using the fact that $0 + x0 = x0$ gives:

$$x0 = 0$$

Using the commutative law $x0 = 0x$ proves that $0x = 0$. □

APPENDIX B—STANDARD RULES OF FRACTIONS

Recall that if $bd \neq 0$, it can be proven from the axioms that $b \neq 0$ and $d \neq 0$, and so both b^{-1} and d^{-1} exist. Further, if $bd \neq 0$, it can be proven from the axioms that $(bd)^{-1} = b^{-1}d^{-1}$. Now suppose that $bd \neq 0$. Then $b \neq 0$ and $d \neq 0$, and so:

$$\begin{aligned} \frac{ad+bc}{bd} &= (ad+bc)(bd)^{-1} \\ &= (bd)^{-1}(ad+bc) \\ &= b^{-1}d^{-1}(ad+bc) \\ &= b^{-1}d^{-1}ad + b^{-1}d^{-1}bc \\ &= ab^{-1}(dd^{-1}) + cd^{-1}(bb^{-1}) \\ &= ab^{-1} + cd^{-1} \\ &= \frac{a}{b} + \frac{c}{d} \end{aligned}$$

which is the derivation of the standard rule $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$ for manipulating fractions.

Other rules such as $\frac{1}{(b/c)} = \frac{c}{b}$ are obtained from the fact $(a^{-1})^{-1} = a$ (which again can be proven from the axioms), provided that a , b , and c are nonzero.

The axioms can also be used to prove that for any real number a , we have $-(-a) = a$ and $-a = (-1)a$. In particular, $1 = -(-1) = (-1)(-1)$.

²Manipulations of adding and/or multiplying an equation by the same thing can be justified from the more basic concept of *substituting into an equation*, but that detail is tangential and is not discussed here.