False Proofs: Where Mathematics Meets Circus Tricks

Rebecca Denise Elkouby

Introduction

When a magician pulls a rabbit from their hat, we marvel at their ability to defy the laws of physics and call the trick "magic". Similarly, when we read a barely ten line proof that seems to obey perfect logic, achieve the absurd conclusion that 1=2, this is another form of circus magic, except that this time it's the laws of mathematics we're defying.

Here is the first such problem I was ever given:

Consider the following proof that 1=2:

Let a, b *be two real numbers such that* $a = b$

Multiply both sides by a: $a^2 = ab$

Subtract b^2 *from both sides:* $a^2 - b^2 = ab - b^2$

Factor both sides: $(a - b)(a + b) = b(a - b)$

Divide both sides by $a - b$: $a + b = b$

Since $a = b$*, replace* a *with* b: $b + b = b \Rightarrow 2b = b$

Divide by b: $2 = 1$

If you've never seen this example before, it can be a little deceiving at first, even though the trick is actually quite simple. When we say "*Divide both sides by* $a - b$: $a + b = b$*"* this is

in fact an invalid manipulation as, given that we've set $a = b$, it follows that $a - b = 0$. Thus we cannot divide by it as division is not defined for 0. The above proof is a rather elementary example of what I'll explore in this essay -

There are many reasons why I love riddles that involve finding where a false proof goes wrong. There's the thrill of the challenge - pinpointing the exact nuance in concepts we think we understand so well. There's the "circus trick" aspect, the way it feels almost like magic, wherein defying the laws of mathematics to achieve 1=2 is almost like defying the laws of physics and having a rabbit pop out of thin air. There's the deeply satisfying feeling of knowing you've mastered a mathematical concept so profoundly that you can spot its limits and the subtle ways in which it can be manipulated. But more than anything, I think the appeal lies in the one constant idea that stands as the backbone to all of these puzzles: mathematics is a perfect system. No matter what our intuition tells us, if a result is wrong, it's not the system that's at fault - it's our reasoning.

There's no other domain in life where the same holds true. In every other field, including the natural sciences, our frameworks are reverse-engineered from observations. Mathematics stands apart: it's a system we define entirely to produce results. The perfection of mathematics lies in this purity of its construction.

In this essay I'll attempt to broadly explore the topic of false proofs in reasonable depth. The first part of my essay will concern itself with some of the necessary formalism required to understand how the logical structures of definitions and proofs are formed in mathematics - and thus how we can exploit them to come up with convincing false proofs. In the second part of this essay I'll consider various different false proofs and divide them through what I've identified to be the structural mechanisms of the "tricks" employed to make us overlook the fallacies in them.

The (many) false proofs I've brought in this essay are in part those I've gathered from friends and teachers over the years and in part thanks to a mathematics stack exchange thread titled "Best Fake Proofs (A M.SE Aprils Fools Day Collection)" that I've attached a link to below. I had a really great time going through them all and highly recommend going over this thread if you enjoy my essay.

The Recipe for Constructing One

What is a mathematical definition?

The very first building block it's important we consider when trying to tackle the topic of a valid or invalid mathematical statement is the definition of the terms we employ in it. I won't be directly discussing definitions in that much depth for this essay, however, a quick exploration into what it rigorously means for something to be "well-defined" is important for understanding the logical flaws embedded in some of the false proofs I'll explore later.

For a mathematical concept to be called "well-defined" it must uphold two properties:

1) existence
2) uniqueness

Starting by understanding the latter, the criteria for **uniqueness** is actually quite simple and intuitive. If I want to define something, not even necessarily in the realm of mathematics, I want my definition to point to something unique. When I tell the grocery store worker, "could you please tell me where I can find the oranges?", I don't want them to point me out to both the "apples" and "oranges". I want the word "oranges" to define a unique object that will usefully help me refer to things in the real world. The same idea holds true in mathematics: when we define a concept, we want that definition to refer to one thing or else we get an inconsistency wherein a pointer refers to two different things. For example, if I were to define x to be the real solution to the equation $x^2 = 4$, x would not be "well-defined" as there are two such solutions (2 and -2).

Existence may initially seem like a bit of a weird criteria for "well - definedness" given that in our day to day use of the word "define" we mean the process of just associating a tag to a "thing" that we can refer to. That is, the "thing's" *existence* doesn't matter so much - we can just semantically assign anything a tag and call that a definition. However, at the risk of getting a bit philosophical for a second, this in essence ignores that in our lived reality, the conceptualization of a "thing" is in and of itself enough to make it exist (if only in the

abstract realm of our conscience). For example, I could define a "unicorn" as a horse with a single horn, and this definition would still be valid in conversation, even if no such creature exists in the physical world as the mere conceptualization (in this case, through assigning it a description) brings the idea to life, at least in a limited abstract sense. Thus the requirement of existence in a definition (of our day to day lives) is trivial which may - falsely - lead us to believe that it's an unnecessary condition.

Unfortunately, mathematics isn't a realm wherein for a thing to exist it can simply float within our brains as an abstraction. Rather it must adhere to all the other pre-existing definitions and rules of the game that we have laid. That is, a mathematical definition must ensure that, for any valid input or case, an entity described by the definition can be constructed or identified. If the definition leads to a situation wherein something doesn't exist for certain inputs, it's not well-defined. For example, defining x_1 to be the integer solution to the equation $x^2 = 3$ would not be a valid definition as there exists no integer x that upholds this equation.

What is a mathematical proof?

The second concept important to understand in order to start talking about a *false proof* is… a proof. That is, let's try to take a second to understand what a mathematical proof looks like and how we can characterize what renders it valid. In this way, we'll be able to identify what's necessary to invalidate a proof and lead us to a false one.

Broadly, a mathematical proof is a rigorous argument that demonstrates the truth of a mathematical statement. It is a logical chain of reasoning that starts from **premises**—statements or facts we assume to be true—and proceeds, step by step, according to the rules of **logic**, until it reaches the **conclusion**, which is the theorem or result we are trying to prove. That is, a mathematical proof consists of:

1) premises
2) logic

Lets understand what both of these are and what they consist of:

Premises are our building blocks for a proof. They consist of *axioms* - fundamental truths that are accepted without proof, they are like the rules of the game of mathematics, we set them as our groundwork and work our way up; *definitions* and *previously proven theorems*.

For example, consider proving the Pythagorean theorem. The premises include:

- The *definition* of a right-angled triangle.
- *Axioms* of Euclidean geometry (e.g., the parallel postulate).
- Any previously proven *theorems* that might be needed, such as the theorem that the sum of angles in a triangle is 180 degrees.

Logic is the framework that binds these premises together. In a way, we can see logic as the current that flows between sound premises together in a way that leads us from point A to point B where point A is all of the things we have thus far established to be true in mathematics (<u>all</u> of our existing premises) and point B is the conclusion of the theorem we are currently considering. There are several types of logical frameworks that can be used in mathematics. To give just a few examples consider:

1. <u>Direct reasoning:</u> This involves moving from premises to a conclusion by a straightforward application of logical rules. A good example for direct reasoning is transitivity which basically says that if (1) leads to (2) and (2) leads to (3) then (1) leads to (3).

2. <u>Proof by contradiction:</u> This involves assuming that the statement we are trying to prove is false and showing that this leads to a contradiction. Therefore, the statement must be true. For example, let's assume that I know that (1) sunflowers and peonies are of two different colours and (2) sunflowers are yellow. We'll use a proof by contradiction to show that peonies are not yellow:

   > *We'll suppose, for the sake of contradiction, that peonies are yellow. Sunflowers and peonies are two different colours, so from this we can derive that sunflowers aren't yellow. Contradiction! Therefore, it must be the case that peonies aren't yellow.*

3. Inductive reasoning: In a proof by induction, we show that (1) a statement is true for an initial case (a natural number) and then (2) prove that if it is true for all natural $k < n$, it must also be true for n. This allows us to conclude that the statement is true for all natural numbers greater than our initial case. The reason for why this works is that we create a sort of "chain of truth" wherein we've proven the base case is true in (1), so from (2) the statement is true for the next number and so on forth. An example for a proof by induction can be found in the next part of my essay.

After all of this somewhat tedious discussion about what mathematical definitions and proofs really are, we're finally ready to answer the question: *so, how do you create a false proof?* Well, if for a proof to be sound it needs true premises **and** valid logic - for a proof to be false, it needs to have false premises **or** invalid logic. What this means is quite simple: our recipe for creating false proofs consists of either including false premises in our proofs - basing ourselves off of things that are mathematically false (say, 1=2), or using invalid logic (say, inferring a conclusion that would only be true in certain cases and generalizing it). Obviously though, here comes the tricky bit.

Psychology of false proofs

Given that mathematics is the most rigorously taught of all subjects and that, on some level - even if only that of intuition, we're all aware of what I explained above (aka the need for true premises and valid logic), the question is begged - *why do we fall for these false mathematical proofs?* Why are we stumped when we see barely ten lines leading to the clearly absurd conclusion that 1=2? Sure, it's a question of involving a false premise or invalid logic somewhere, but we obviously wouldn't fall for a 1=2 appearing in the middle of a proof right? Herein lies the true art of false proofs - hiding and masquerading what in another light would be to us irrefutably false through clever manipulations - whether they be semantic, algebraic or other. In this section of the essay I'll consider a couple of what I've identified to be the systematic causes for why we let slip false mathematical manipulations

and I'll draw out some broad mechanisms to categorize false proofs based on these cover-up techniques.

Wrongly taught

I'd like to start with what I see as the simplest, most obvious, and also probably most avoidable reason for why we fall for false proofs. Put plainly, sometimes maths is just badly (or insufficiently) taught and that leads us to let things fall through the cracks. I'd like to look at what i consider to be a really simple and deceptive example that encompasses a lot of the trickiness embedded in the task of teaching maths sufficiently well for it to not lead to logical inconsistencies, but at the same time be accessible enough that seventeen year olds can understand it.

Consider the following (false) proof that $1 = -1$:

$$1 = \sqrt{1} = \sqrt{(-1) \cdot (-1)} = \sqrt{-1} \cdot \sqrt{-1} = i^2 = -1$$

To the eye of the unsuspecting high school student who was taught that $i = \sqrt{-1}$, nothing here seems problematic and it is in fact unclear what in this (tiny!) proof that leads to the very obviously wrong conclusion that 1=-1 is off. Maybe some would answer that it is that the arithmetic manipulation of multiplicity that says that $\sqrt{ab} = \sqrt{a} \cdot \sqrt{b}$ does not hold for the realm of complex numbers, but that seems a little bit like cheating. It doesn't provide any deep reason as to why the reasoning is wrong and if anything is more consequence than cause. That is, this rule does not hold for the more inherent reason that we have yet to discover, not the other way around.

Another important note that I'd like to make here is that if we notice that a certain rule does not hold for a mathematical function when we expand its scope of definition (so as we did here by expanding the square root as an operation that could function on negative numbers in addition to non-negative ones), thats an indication that we cannot expand the definition. The reason here is quite simple. If we have a mathematical function, we want it to be consistent.

Otherwise we can't really do much with it, we can't mix real and complex numbers as they dont have the same properties for that function, and in general things get quite messy. It is not necessarily the case that properties of functions must be consistent over different fields, but it would definitely make sense for that to be the case and thus should be a good intuition check for the validity of our definition.

Hopefully, the feeling that you're starting to get by this point is that there's something off with the well-known definition of i as $\sqrt{-1}$.

Let's remember how the square root function is defined:

the square root of a number a is the positive solution to the equation $x^2 = a$.

That is, seemingly, i would be the *positive* solution to the equation $x^2 = -1$ (remember from above that it cannot just be "the solution to the equation" as there would be two such possibilities and the square root would not be well-defined, thus the positivity requirement is necessary). But what does it mean for a number to be positive? This is a little beyond the scope of my essay, but put briefly, for a number to be positive it must belong to an ordered field (a special type of set) that is ordered for some order relation '>'. I wont get into the definition of an ordered field, or an order relation, but one of the notable properties we can derive from these definitions is that all numbers $x \neq 0$ in an ordered field must obey $x^{2'} > '0$ for our special "order relation". However, $i^2 = -1$ and using our definitions for an ordered field we can also prove that $0' > ' - 1$, no matter what our field is. Thus, the field of complex numbers cannot be an ordered field. This leads us to the inevitable conclusion that there can be no notion of "positivity" in the field of complex numbers. From this problematic initial setup we'll be able to derive countless false results, as in fact, the classic definition of i as the square root of -1 is not well defined. Notice also that it makes perfect sense that one of these such results is of the form "$1 = -1$" as the problem in our definition was one of lack of uniqueness (there is no meaningful way to distinguish between i and $-i$.)

Lack of Rooting in Reality

You (the reader) and I (the writer) can all clearly form strong intuitions around the concepts of integers or euclidean geometry because we can translate $n \in N$ to a reality of having n apples or an equilateral triangle to drawing three connected lines of equal length on a page. The same, however, cannot be said for certain mathematical concepts such as infinity, higher-dimensions and complex numbers that have no (or no obvious) rooting in our lived reality. Herein lies the second of what I've identified to be the prominent reasons for why we fall for false proofs. That is, our struggle to comprehend and relate to abstract concepts that are not rooted in our tangible reality. This detachment can lead to misunderstandings and the acceptance of incorrect conclusions, as we may not fully grasp the implications or nuances of these concepts.

On a holistic level, the reason for why we have a harder time catching fallacies and false manipulations when dealing with abstract ideas is that we can't rely on our intuition to follow and "fact-check" proofs when we read them and therefore we can only rely on the abstract rules of the abstract concepts. This means that our only way of detecting a false manipulation in the steps of the proof is through rigorously making sure each step of the proof adheres to the definition of the concepts we're dealing with, and not through a broader understanding of why something may be wrong.

To give an illustration, consider a proof wherein I (falsely) divide a number by zero. I have a basic intuition of what division and zero are so it's likely that I look back and feel something is wrong. The reason I can do this is because through applying my understanding that dividing m by n means taking m objects and dividing them to n people, trying to divide m objects to 0 people immediately seems weird. I could say each person gets 0 objects because there are 0 people but through that same principle, I could give any amount of the m objects to "each person" and it would be the same. Thus, without even the rigorous definition of division and the principle of not dividing by 0, I'm able to detect that something's wrong through my understanding that a definition must be unique. The problem arises when the same process of using our intuition of concepts to "fact-check" a proof can't really be used because we have no tangible intuition of the concepts we're dealing with.

Let's get a little more concrete. Consider the following proof that $\sum_{n=0}^{\infty}(-1)^n = \frac{1}{2}$ that I'm unfortunately certain many of you have seen before:

We'll denote $S = 1 - 1 + 1 - 1 + \ldots$

We can notice that: $S = 1 - S$

Therefore: $2S = 1 \Rightarrow S = \frac{1}{2}$

That is: $1 - 1 + 1 - 1 + \ldots = \frac{1}{2}$

However, this is obviously wrong given that this series just oscillates in the values it takes ($\sum_{n=0}^{k}(-1)^n$ is 0 when k is odd and 1 when k is even) and doesn't even converge - it being equal to $\frac{1}{2}$ is clearly absurd.

I know that there exists countless "proofs" of this theorem on the internet, along with ones of the even more absurd result $\sum_{n=1}^{\infty} n = 1 + 2 + 3 + \ldots = \frac{-1}{12}$ that try to convince us that this is in fact true, and specifically they cash in on the fact that the notion of infinity and infinite sums is beyond the grasp of our intuition to convince us that even if it seems unreasonable, "the math says it, so it's true".

Where this "proof" goes wrong is in fact quite simple, it assumes the existence of S (the sum of the series) without proving it exists. That is, it assumes there exists a finite number $S \in R$ such that $1 - 1 + 1 - 1 + \ldots = S$, whereas this is in fact not only not trivial, it's completely wrong. So, sure, *were* this S to exist then all the manipulations on it are valid and it would in fact have to be equal to $\frac{1}{2}$ but it existing is a huge "if." In fact, we can be certain that it doesn't exist as such an S would defined to be

$\lim_{k \to \infty} \sum_{i=0}^{k} (-1)^k = \lim_{k \to \infty} 0\,\{if\ k\ is\ odd\},\ 1\,\{if\ k\ is\ even\}$ which is clearly a limit that does not exist. The reason we're so confused by this is that in finite sums we don't need to prove existence for sums, it's implicit. Herein comes to play the idea of our lack of ability to intuition check when it comes to abstract ideas in mathematics, because infinity is such an unconcrete concept idea to us - the closest we have when it comes to forming intuitions about infinite sums is how we work with finite ones - which turns out to be completely misleading.

To illustrate just how non-trivial the idea of existence of a result is, I'll bring to you another false proof from a completely different context that has the exact same fallacy as the one above. We'll prove that 1 is the greatest natural number (I don't think there'll be much controversy in saying that this one is definitely false):

Let n *be the greatest natural number.*

Given this property of n*, it is necessarily the case that* $n^2 \leq n$.

Thus $n(n-1) \leq 0$.

We can thus conclude that $0 \leq n \leq 1$

n *is natural so* $n = 1$

Hopefully this proof makes it clear why we must prove the existence of a solution before claiming to find it. In many cases this requirement is trivial (for example, a finite sum always sums up to a finite number - we just need to find it, not prove it exists) which is why we can be led to think it's not a necessary step of finding a solution.

But the above proof for the oscillating sum is not an outlier. To illustrate how incredibly unreliable our intuition can be when it comes to the concept of infinity and infinite sums, I'd like to present two really interesting results:

1) This first is a theorem called the Riemann Rearrangement Theorem that says the following: We'll define a series to be "conditionally convergent" if it converges, but the series of its absolute values diverges.

Put mathematically we'll define $\sum_{i=0}^{\infty} a_i$ as "conditionally convergent" if $\sum_{i=0}^{\infty} a_i$ converges (it has a limit) but $\sum_{i=0}^{\infty} |a_i|$ diverges (it goes towards infinity).

The theorem states that given any conditionally convergent series, for all real numbers x, there exists a rearranged order of the series such that the rearranged series converges to x. Let's consider a simple example to unpack this theorem:

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \ldots$$

This series is conditionally convergent because whilst the series itself converges to $ln(2)$, the series of absolute values diverges (I won't include the proof here but I invite you to look this up in the below resources on the harmonic series).

Now, suppose we want the series to sum to a different number, say 3. To achieve this, we can rearrange the terms so that the positive terms dominate initially, pushing the partial sums closer to 3, and then gradually include enough negative terms to stabilize the total at 3. Specifically, we could group the terms to first sum many positive contributions until the partial sum exceeds 3, and then balance it by adding enough negative terms to bring it back toward 3. By carefully repeating this process, the rearranged series will converge to 3.

The key idea in this theorem is that the positive terms and negative terms each "pull" the sum in opposite directions. Since the series is conditionally convergent (its absolute values diverge), you can exploit this property to "steer" the total sum toward any target that you choose.

The bottom line here is that the commutative property of addition that we're so used to that says that $a + b = b + a$ (that is, we are allowed to reorder the elements in a sum) does not hold true for infinite sums!

2) The second thing I'd like to consider in terms of how unreliable our intuition is when it comes to infinite sums is the idea of brackets in a sum. We're all familiar with

associative property of addition that says that $(a + b) + c = a + (b + c)$. In fact, it's so elementary to us that you probably didn't even think of this as a "special property" of real numbers before, it's just a manipulation you do naturally. However, let's consider the above series: $1 - 1 + 1 - 1 + \ldots$ and look at two possible bracket placements:

I. $(1 - 1) + (1 - 1) + \ldots$

II. $1 - (1 - 1) - (1 - 1) - \ldots$

both of these placements are seemingly equivalent to our original series $\sum_{n=1}^{\infty} (-1)^n$ but notice that I. = 0 whereas II. = 1! The necessary conclusion here is that when it comes to infinite sums we cannot just naively apply associativity as we would with finite sums, even though our intuition begs us to believe that the order of addition (this is what's being changed under bracket placement) should have no effect on our result.

To tie all this back to our discussion on false proofs, notice just how easily we could have misused either one of these properties to create all sorts of absurd results. The examples above showcase just how fragile our intuition can be when we extend mathematical properties from the finite to the infinite. What feels like a natural extension of basic rules - be it the commutativity of addition or the reliability of grouping terms - can break down in surprising and counterintuitive ways when faced with infinity. This dissonance between our finite-based intuition and the abstract, unconcrete nature of infinity lies at the heart of why so many false proofs feel convincing.

A General Rule With Exceptions

Let's consider the following (obviously false) theorem: All people have the same eye color.

Proof by induction:

The statement "All members of any set of people have the same eye color" is clearly true for any empty set (if there are no elements in a set, any statement is true for all of those elements regardless of what it is).

Now, assume we have a set S of people, and the inductive hypothesis is true for all smaller sets. Choose an ordering on the set, and let S_1 *be the set formed by removing the first person in the ordered set, and* S_2 *be the set formed by removing the last person in the ordered set.*

All members of S_1 *have the same eye color (by the induction hypothesis), we'll call it* x*, and so do those of* S_2*, we'll call this color* y*. However,* $S_1 \cap S_2$ *has members from both sets, we'll choose one of these members. Its eye color is both x and y and as a person has a unique eye color,* $x = y$*. Thus, all members of S must have the same eye color.* □

The problem with this proof lies within the phrase: "However, $S_1 \cap S_2$ has members from both sets, we'll choose one of these members". Let's think about it a little more deeply. Whilst for most sets this seems like a reasonable statement - taking off one member from two common sets should leave common ground, this is very much not the case for our edge cases. That is, for all sets with at least three elements it is true that taking two subsets of that set, each being the set minus one member, the subsets contain at least one member in common as there are enough members that removing one leaves common ground - for sets with one or two elements this won't be the case. Consider a set with one element. By any ordering the first and last element will be the same and therefore the subsets S_1 and S_2 will both be empty sets - such that they have no common member. Herein lies the flaw that breaks the logical chain of the above argument - because an inductive proof works as a sort of chain wherein the truth of the statement at each stage requires the truth of the links beforehand, this odd case to the rule for sets with one or two elements is enough to render the claim void for all non empty sets.

We can visualize inductive proofs a little bit like a tall tower of thin blocks stacked one on top of the other (with block n representing the truth of the hypothesis for n) such that when we pull out one block - all those on top of it will consequently fall out. What this means is that

the combination of a rule with an exception and proof by induction is pretty much a lethal recipe for a false proof. Whilst its true that many times edge cases that are false wherein a general rule is true aren't usually too problematic and we tend to forget them, when we're using proof by induction, the proof of each consequent stage is dependent on the one before such that its enough for a single such "edge case" to be false for our entire theorem to be false.

Another really great example for a general rule with exceptions being a good way to form false proofs is the false proof I brought in the intro that "shows" that 1=2. We saw that the logical error in this proof was that when we divided both sides of our equation by $a - b$ this was in fact invalid as $a - b$ was equal to 0. The reason that we were able to get away with this at first glance is that dividing both sides of an equation by a number n is allowed for all values of n except for 0. That is, we have a rule that is true almost always, infinitely so, except for in one case. This can make us careless in applying the rule, forgetting that there exists that one case wherein that thing that seems so trivially valid is in fact completely problematic.

Notation

Let's look at what I consider to be quite a neat false proof claiming that 0=1.

Proof:

Let's consider the integral $\int \frac{1}{xlogx} dx$

we can integrate by parts where u'=$\frac{1}{x}$ *and v=*$\frac{1}{logx}$ *so u=*$logx$ *and v'=*$\frac{-1}{x \cdot log^2 x}$*:*

the formula for integration by parts is: $\int u'vdx = uv - \int uv'dx$

that is: $\int \frac{1}{xlogx} dx = 1 + \int \frac{1}{xlogx} dx$

we'll subtract $\int \frac{1}{xlogx} dx$ *from both sides and we'll be left with: 0=1* □

What's wrong with this proof actually just lies within the very last line, "*we'll subtract* $\int \frac{1}{xlogx} dx$ *from both sides"*. To understand why, let's remember a few important things about indefinite integrals. Before I get into a formal rundown of what this proof gets wrong, we'll form a good intuition. As every high school maths teacher has certainly ceaselessly drilled in their students minds, when we write out an indefinite integral we must always remember the constant representing the fact that an integral is insensitive to a displacement of a function along the y axis. That is $\int xdx = \frac{1}{2}x^2 + C$ and not just $\frac{1}{2}x^2$. What this therefore means for our previous equation is that $1 + C_1 = C_2$ where C_1 and C_2 are the constants of both integrals respectively. (Note, that's not to say that C_1 and C_2 represent two numbers whose difference is 1 and they're the fixed constants of the integrals. I just wanted to illustrate that the constant can be different for both, thus why we get a difference of 1. What I'm saying here is messy, but trust me it's just to form an intuition, I'll get to the rigorous rundown in a second.) Hopefully you can start to see from here that because integrals can be differentiated up to a constant, the fact that we get a "+1" isn't too problematic as 1 is in and of itself a constant.

That being said, all our steps in the proof were quite rigorous, why would subtracting cause a problem? To understand this, let's get a little more formal.

The reason why we can't subtract the integral from both sides is that the subtraction operation treats the indefinite integral as an algebraic quantity that can be canceled out, like a number or a variable. This assumption fails to consider that integrals represent families of functions differing by constants, or put more formally - sets. So, taking into consideration the uniqueness theorem for integrals that states that the primitive functions of a continuous function can only differ by a constant, $\int \frac{1}{xlogx} dx$ is actually equivalent to $S = \{g(x) + c \mid c \in R\}$ where g is some primitive function of $f(x) = \frac{1}{xlogx}$. Once we've established this, let's reconsider the right hand side of the equation. We have $1 + \int \frac{1}{xlogx} dx = 1 + S$. But what does $1 + S$ mean? How can we add a number to a set

given that they're two different algebraic structures? The answer here is obviously that we can't. This $1 + S$ is once again our way of simplifying the story. If we think about what we're trying to say, we want 1 to be added to each one of the primitive functions S contains. Thus, $1 + S = \{g(x) + c + 1 \mid c \in R\}$. However, that's the exact same thing as $\{g(x) + c \mid c \in R\}$, given that both simply take g and add to it every real constant possible.

Thus, $\int \frac{1}{xlogx} dx = 1 + \int \frac{1}{xlogx} dx$ is not contradictory at all, in fact, it's entirely trivial!

Our source of confusion stems from our notation which aims to simplify things for us but ends up letting loose through the cracks the real nature of an indefinite integral as a set, thus making us think we can simply "subtract" it from both sides as we would with numbers.

Let's look at another example wherein notation is misused to lead us to a false conclusion. Consider the following proof that 0=1:

Notice that: $x = 1 + 1 + \ldots + 1 \{x\ times\}$

We'll take the derivative: $1 = \frac{d}{dx} x = \frac{d}{dx} (1 + 1 + \ldots + 1) = 0 + 0 + \ldots + 0 = 0$

□

This is such a classic example of how misleading our writing can be in maths and underlines really clearly how deceptive notation can easily gloss over the true meaning of a term. There are actually two big problems within this proof.

The first lies in the statement *"*$x = 1 + 1 + \ldots + 1 \{x\ times\}$*"*. Notice that this doesn't even really make sense. We define addition of n terms for a positive whole n but what would it even mean to add x terms (in this case x 1s) if x isn't a whole number? Or not even rational? In this way the notation of x as a sum of x ones is inherently faulty because addition is defined only for discrete numbers. For the concept of "non whole number addition" we have multiplication and the intuition may certainly be adding a set of units a certain amount of times but this $x = 1 + 1 + \ldots + 1 \{x\ times\}$ is certainly wrong. Given that derivatives are taken only over derivable functions that must be continuous (intuitively, we can't find the

derivative of a function with "holes in it") the function $f(x) = x$ that we're taking the derivative of here must take on non-whole values and thus the first statement is wrong.

The second problem within this proof is simply that it's doing the derivative wrong. It uses the property of additivity for derivation that says that if $f(x) = g(x) + h(x)$ then $f'(x) = g'(x) + h'(x)$ but uses this rule incorrectly as it is only true for a constant number of additions (and here the number of additions is 'x' - a function and not a constant). We'll remember that the definition of a derivative at a point is $f'(x_0) = \lim_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0}$. Thus, even if we were to accept the notation we just saw was problematic we'd get: $f'(x) = \lim_{x \to x_0} \frac{1+\ldots+1\{x\ times\} - 1+\ldots+1\ \{x_0\ times\}}{x-x_0} = \lim_{x \to x_0} \frac{1+\ldots+1\{x-x_0\ times\}}{x-x_0} = 1.$

If we consider what allowed us to overlook these two slights, it's probably the notation of how the proof was presented. We're used to adding an unknown number of numbers using the _ +… + _ notation and thus we ignore the fact that it requires the addition to be for a discrete sum. Additionally, the additivity property of the derivative requires that it be for a finite number, which is not the case for this "addition", and yet we overlook it as the notation allows us to forget how many terms are being added: it's a constant format of "number" plus "numbers" plus "number" and so we're misled into thinking we can use the well known manipulation for derivatives.

Ambiguous Definition

Writing this essay helped me understand that maths has a rather big problem when it comes to its language. Inevitably, because it's a game (or an art, or a science - interesting topic of discussion but probably too long for this essay) employed by humans - its language has to **be human language** or to **become human language**. Phrases like "There's a 50-50 *chance*", "That idea is *derivative*", "it was a *random* occurrence" or the "the *average* person" all use inherently mathematical concepts for the description of day-to-day occurrences. The problem here is that not all humans are mathematicians and so inevitably some of these concepts

won’t be used in their correct way, especially given the fact that they have complex nuances (so for example, the *average* individual usually doesn't give much thought to the fact that to talk of an average is to first identify a set, decide on a metric of comparison, calculate each person's score on that scale and find the mean).

But even if all humans were mathematicians, life is too dynamic, too nuanced and requires too many levels of abstraction to ever be a safe playing ground for mathematical terms if we wish to keep them intact. Unfortunately, humans are rather bad at sticking things to their specific domains and so if we see something in life that reminds us of that thing that we defined as *randomness* - we’ll call it *random* even if it's only just very hard to predict. The bottom line here is that because it is humans who do maths, its language gets intertwined with our general lives and this inevitably leads our desirably rigorous mathematical terms to get riddled with ambiguous definitions. I’ll give two examples of this below and consider how it effectively allows us to create false proofs.

First, let's consider a proof to the theorem that says that a dog has 9 legs:

> *No dog has 5 legs,*
>
> *A dog has 4 more legs than* ***no dog****.*
>
> *Thus, a dog has 9 legs* □

This false proof is rather simple and it probably didn't take you too long to figure out what the problem with it is. However, it underlines a really critical idea. Here, the string of words “no dog” is being misused by referring to two completely different things (validly, within the context of the English language). The first time no dog appears its meaning is “none of the dogs” whereas the second appearance of the phrase **no dog** refers to “an absence of dogs”. These are inherently two different meanings and indeed, if you try and switch up the phrases you’ll get them to mean - “an absence of dogs has 5 legs” and “a dog has 4 more legs than none of the dogs” - both sentences that barely make sense and definitely aren't true. Thus, each of the statements “no dog has 5 legs” and “a dog has 4 more legs than no dog” are individually true but because both “no dog”s aren't the same, we cannot apply the logical

structure of transitivity that's currently being used to reach the conclusion “A dog has 9 legs”. (If you’re acute you’ll probably also have noticed that the phrase “a dog” is also somewhat problematically defined in the above proof). The idea this false proof employs

The final false proof I'll bring in this essay is by far the most interesting of all the ones I've considered whilst researching this topic. I’ll admit that I had a hard time pinpointing why exactly it was wrong and I’ll do my best to explain this point given my acquired understanding, but in case it's still not clear by the end of my explanation, I invite you to look up the resources I attached below about the Berry Paradox that personally helped me understand it a lot better.

Let's consider the following proposition: all natural numbers are definable in under eleven words. Now, this proposition is clearly wrong. The reason here is that there are infinitely many natural numbers and finitely many words in the english language - let’s call this number x - so there are $\sum_{i=1}^{10} x^i$ phrases with under 11 words in english. Thus, given that each phrase can describe at most one number (remember that for something to be well defined, it must be unique), it is impossible to define infinitely many numbers in under 11 words.

However, consider the following proof by contradiction for the above proposition:

> *Suppose for the sake of contradiction that not all positive natural numbers are definable in under eleven words. Then there is a smallest integer n∈N which is not definable in under eleven words. But this number is*
>
> *the smallest positive integer not definable in under eleven words,*
>
> *Therefore, it is definable in ten words. Contradiction!* □

This contradiction is known as the Berry Paradox and underlines the idea of ambiguous definitions in a really great way. Let’s dive into it to gain an understanding of where it goes wrong.

In natural languages (languages we use to communicate between one another, like English, French, Hebrew etc.) we can create a distinction between two different types of ways to attempt to define something:

1) **Semantic Definitions:**
   These are definitions that attempt to define an object based on the meaning of that object. For example, defining a pen as “a cylindrical tool containing ink used to write on a page” would be a semantic definition of a pen as it attempts to define it through its properties.

2) **Syntactic definitions:**
   These are meta-definitions that use properties of expressions, such as the number of words in a phrase, alphabetic combinations and more to define an object. An example for this would be as in the paradox “the smallest positive integer not definable in under eleven words” as we are defining this number not through its mathematical properties, but rather through linguistic properties.

Within the context of natural languages these are both valid ways to define and object as we would be able to understand what the definition was referring to in both cases. However, in formal systems, this is problematic because natural languages allow expressions (namely, syntactic definitions) to serve two roles: they act as both objects *within* the language and syntax *defining* the language. This lack of separation creates self-referential ambiguity that makes it such that terms are not well-defined.

That is, the problem within the above proof lies in this definition which is simply not valid in the context of mathematics. A phrase like this cannot be a definition in mathematics as in a formal system expressions cannot simultaneously be objects of the language, as well as valid syntax of the language. The act of defining “the smallest positive integer not definable in under eleven words” **makes** the definition itself become part of the system it is trying to evaluate.

If you’d like to picture a little more concretely what in this definition is problematic, I invite you to think of it like a recursive code. When we say, "the smallest positive integer not definable in under eleven words," let’s focus on the word “definable.” To say something is

“definable” means that a valid definition for it exists. So, to evaluate this phrase, we need to examine the set of “all possible definitions”. However, the phrase itself - "the smallest positive integer not definable in under eleven words" - is a definition. This means it is a member of the set we’re evaluating. Thus we’re led back to it and we need it to be defined in order to define it! We’ll zoom into it again (as is necessary to understand what it refers to) and go through this process indefinitely in a way that creates an infinite recursion.

If we want to be rigorous and tie this back to our discussion on valid and invalid mathematical definitions we can identify the criteria missing here as that of existence. This is like writing a recursive function without a base case - it calls itself indefinitely without ever resolving to a concrete value, we never ground into an actual object, thus we’re defining something that does not exist.

Conclusion

In the introduction to this essay I reflected on one of the aspects to what makes false proofs so captivating: the fact that mathematics is a perfect system. False proofs thrive on this premise - they challenge us to identify where our own logic has faltered, providing a thrilling puzzle whose solution reinforces the perfection of mathematics.

But is mathematics truly perfect? Throughout this essay, I’ve explored how false proofs expose subtleties in *our* logic, language, and intuition, but maybe there’s something incomplete about mathematics itself? While some false proofs stem from simple missteps, others point to something deeper - mathematics’ own limitations.

I’d like to end this essay with one final false proof, probably the coolest “circus trick” I hold in my arsenal that I'll flourish for my grand finale. This proof will also help us consider the limitations of mathematics as a system.

Let's look at the following proof of Riemman’s hypothesis:

> *Consider the following three statements:*
>
> > *(1) At least one of the following statements is true*

(2) The above statement is false

(3) Riemann's hypothesis is true

If (1) is true, then (2) cannot be true and thus, by (1), (3) is true.
Else, (1) is false and thus none of the below statements are true (otherwise (1) would be true) and thus (2) is false ⇒ (1) is true. Contradiction!

Thus, it must be the case that (1) is true and Riemann's hypothesis is true!

Before you get all excited, unfortunately, no, we did not just prove one of the greatest unsolved theorems in mathematics today. What's going on here is an exemplification of **Tarski's Undefinability Theorem**, which states that the truth of a statement cannot be defined within the same system that expresses it. Simply put, a statement like "This statement is false" is not valid in mathematics and will necessarily result in a contradiction. In the context of this proof: statement (1) tries to "define" the truth of (2) and (3), but (2) refers back to (1), making the system self-referential. This circular dependency leads to an inability to assign consistent truth values, illustrating the fundamental limitations of expressing "truth" within a system.

(Maybe you'll have noticed that the fallacy in Berry's Paradox I explained above boils down to exactly the same idea expressed here. I won't go into this in depth but I invite you to ponder this idea.)

Gödel's incompleteness theorems and Tarski's undefinability theorem deepen this insight. Gödel showed that no sufficiently powerful system can prove all truths about itself, and Tarski revealed that truth itself cannot be fully captured within the system. Together, they demonstrate that mathematics, for all its elegance and rigor, has boundaries it cannot surpass.

The false proof of the Riemann Hypothesis exemplifies this beautifully. In essence, the problem in this false proof is not really due to our own limitations and misunderstanding, but rather fundamentally because mathematics is not a perfect system. Its flaw exposes how self-referential statements disrupt the illusion of a complete, self-contained system. In this

way, mathematics mirrors the paradoxes of human thought - it is both a tool of unparalleled precision and a discipline constrained by the limitations of logic and language.

I'll let you (the reader) choose on what note you leave my essay. Maybe you'll have read it as an enjoyable presentation of the "behind the scenes" of a magician's impressive tricks. Maybe you'll have read it as an exploration of the dichotomy of imperfectness, the back and forth between where our own logic fails and why, and where the system fails and why. In either case, I hope you've enjoyed the show!

Bibliography

https://www.jamesrmeyer.com/paradoxes/berry-paradox

https://math.stackexchange.com/questions/348198/best-fake-proofs-a-m-se-april-fools-day-collection

https://en.wikipedia.org/wiki/Berry_paradox

https://en.wikipedia.org/wiki/Ordered_field

https://en.wikipedia.org/wiki/Mathematical_proof