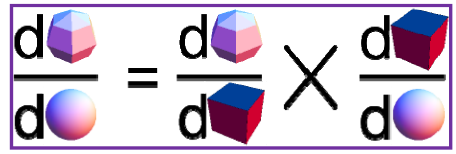
# Toward a Proof of the Chain Rule

**Joe Gerhardinger**, jgerhardinger@nda.org , Notre Dame Academy, Toledo, OH

**Abstract**

The proof of the chain rule from calculus is usually omitted from a beginning calculus course. Six different rationales and motivations are presented here to help students understand some of the rule's details and the reason the rule is valid.



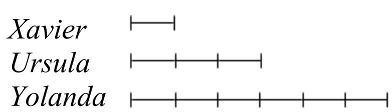
**The chain rule is the most commonly used rule of differentiation, but most calculus texts include only one intuitive justification.** The official proof, which contains a somewhat technical complication, is often relegated to the appendix. Teachers may not be familiar with a full complement of (six) different rationales, including multiple representations and analysis of special cases, which can be arranged to culminate in deeper student understanding.

## The Liebniz Version

Students already have an intuition about how the chain rule works. Here's a situation about three runners:

*Yolanda runs twice as fast as Ursula.*
*Ursula runs three times as fast as Xavier.*
*How much faster does Yolanda run than Xavier?*

Students readily answer that Yolanda runs six times faster than Xavier, and they used the chain rule without realizing it. When pressed to explain why this works, students draw a picture of the situation, showing the runners' relative units of distance traveled per unit of time.

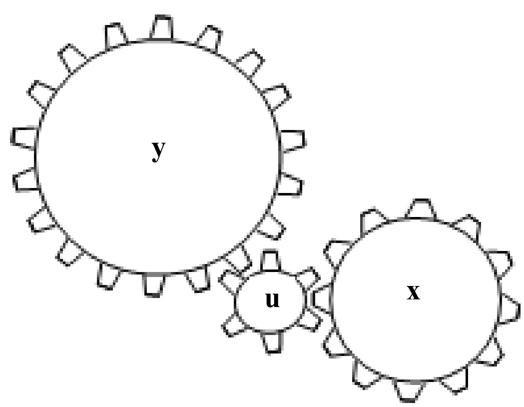



Since rate is distance over time, this information can be represented with derivatives.

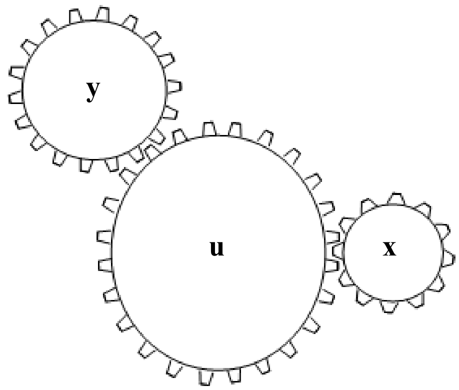
| | |
|---|---|
| *Yolanda runs twice as fast as Ursula.* | $\frac{dy}{du} = 2$ |
| *Ursula runs three times as fast as Xavier.* | $\frac{du}{dx} = 3$ |
| *How much faster does Yolanda run than Xavier?* | $2 \cdot 3 = 6$, so $\frac{dy}{du} \cdot \frac{du}{dx} = \frac{dy}{dx}$ |

This is the first time my students see the chain rule, and the students *themselves* discover it.

Another useful motivation of the Liebniz chain rule involves a series of interlocking gears (Finney, et al., p. 141). Students can count the teeth in the diagram, finding that **x** has 12 teeth, **u** has 6 teeth, followed by **y** with 18. Thus, they reason that for every one rotation of **x**, **u** will spin twice, and for every one rotation of **u**, **y** spins 1/3 times. In Leibniz notation, this means that $\frac{du}{dx} = 2$ and $\frac{dy}{du} = \frac{1}{3}$. Students can determine how many times **y** spins for every one spin of **x** by visualizing the gears turning. They answer $\frac{2}{3} = 2 \cdot \frac{1}{3}$, which in Liebniz notation is $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$ .



Some functions can be written as composites in different ways and, to confirm that this will not change the final derivative, students can examine what happens when the size of gear **u** is changed. For instance, if **u** has 24 gears, then $\frac{du}{dx} = \frac{1}{2}$ and $\frac{dy}{du} = \frac{4}{3}$.




Finding $\frac{dy}{dx}$, and thus confirming the chain rule, can again be done by visualizing the gears turning. This time, least common multiples are helpful: every 2 rotations of **x** yield 1 rotation of **u**, so 6 rotations of **x** give 3 rotations of **u**, which give 4 rotations of **y**. Thus, $\frac{dy}{dx} = \frac{4}{6} = \frac{2}{3}$. As before, $\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$.

It looks as though the chain rule would be easy to prove, but the Liebniz symbol for the derivative is just notation, and the numerators and denominators cannot cancel.

### The Function-Composition Version

The function version of the chain rule, $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$, is more difficult for students to use and may remain more mysterious to them unless motivated separately.

#### *A Special Case*

In the special linear case of an inner function, the concept of horizontal transformations can be used to help motivate the rule. I developed this technique over several tries, with encouraging results. The three types shown correspond to the three horizontal transformations which students know well. I use a parabola for the original parent function graph because of students' familiarity with it.

Horizontal Stretch/Shrink:

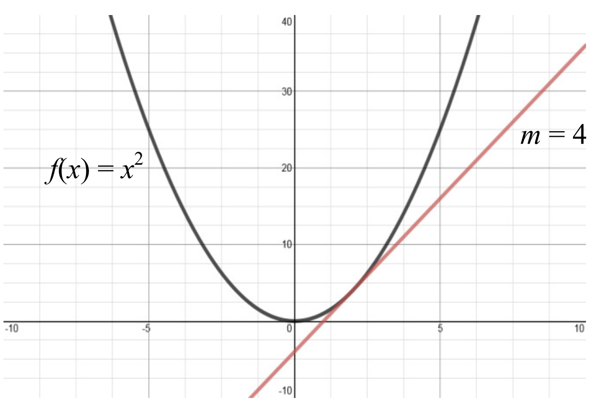



Figure 1. original graph: $f(x) = x^2$, tangent line at $x = 2$

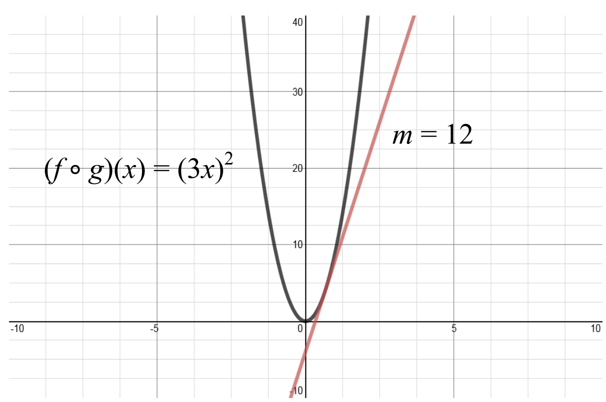



Figure 2. horizontally shrunk by 3 $(f \circ g)(x) = (3x)^2$, tangent line at $x = 2/3$

In Figure 1, we start with the graph of $f(x) = x^2$, and students know that $f'(2) = 4$. In Figure 2, we've taken the original parabola and its tangent line and horizontally shrunk both by 3. Students can predict the changes which take place. The new line is tangent at $x = 2/3$ (closer to the $y$-axis by a factor of 3) and its slope is 12 (because the line is 3 times as steep as the original tangent line). Both changes appear in the formulation of the chain rule as we use it to evaluate the slope of the transformed tangent line (at $x = 2/3$): one takes the slope of the original tangent line (at $x = 2$) and multiplies by the transformation factor (3 in this case).

Letting $f(x) = x^2$, $g(x) = 3x$, and $g'(x) = 3$:

chain rule: $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$

at $x = 2/3$: $(f \circ g)'(2/3) = f'(g(2/3)) \cdot g'(2/3)$

$(f \circ g)'(2/3) = f'(2) \cdot 3$

$12 = 4 \cdot 3$

↑ transformed slope (12), ↑ original slope (4), ↑ transformation factor (3)

This example helps clarify several ideas. Of first importance, it illustrates why the chain rule involves a product, because the transformed slope naturally involves both the original slope and the transformation factor. Next, I ask students about the transformation factor itself. I call it the *transformation factor,* initially, in order to make use of their intuition about transformations. But the transformation factor is just actually just the slope of $g$, the inner function. Thus, they now have the understanding that the chain rule involves the product of the slopes, which are the derivatives of both the outer function $f$ and inner function $g$. Finally, this illustrates why the derivative of $f$ is evaluated at $g(x)$, as it is how to obtain the slope of the original tangent. The other two horizontal transformations help reinforce these points.

Horizontal Flip:

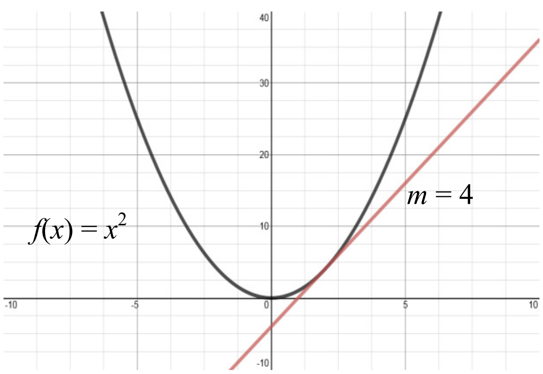



Figure 3. original graph: $f(x) = x^2$, tangent line at $x = 2$

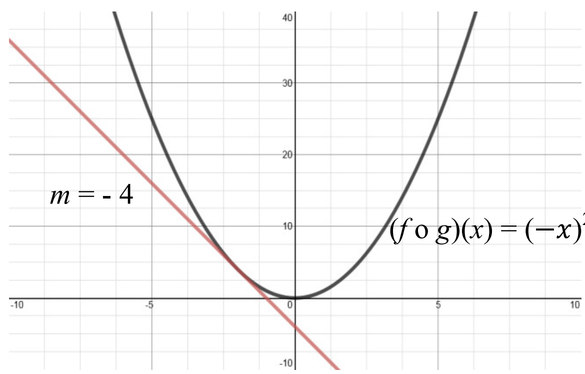



Figure 4. horizontally flipped $(f \circ g)(x) = (-x)^2 = x^2$ tangent line at $x = -2$

In Figure 3, we start the same way as before, with the graph of $f(x) = x^2$ and its tangent line at $x = 2$ with slope of 4. In Figure 4, we've taken the original parabola and its tangent line and horizontally flipped both over the $y$-axis. The new line is tangent at $x = -2$ and its slope is $-4$. To find the slope of the transformed tangent line (at $x = -2$), one takes the slope of the original tangent line (at $x = 2$) and multiplies by the transformation factor ($-1$, in this case).

Letting $f(x) = x^2$, $g(x) = -x$, and $g'(x) = -1$:

chain rule: $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$

at $x = -2$: $(f \circ g)'(-2) = f'(g(-2)) \cdot g'(-2)$

$(f \circ g)'(-2) = f'(2) \cdot -1$

$-4 = 4 \cdot -1$

(transformed slope) = (original slope) · (transformation factor)

Again, this shows that the chain rule involves the product of the slopes of both the inner and outer functions, and that the derivative of *f* is evaluated at at *g*(*x*), the original tangent line location.

### Horizontal Shift:

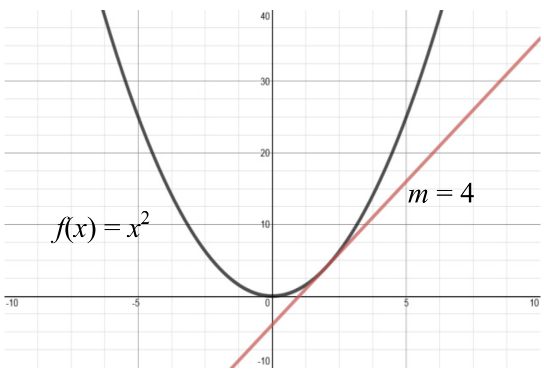



Figure 5: original graph:
$f(x) = x^2$,
tangent line at $x = 2$

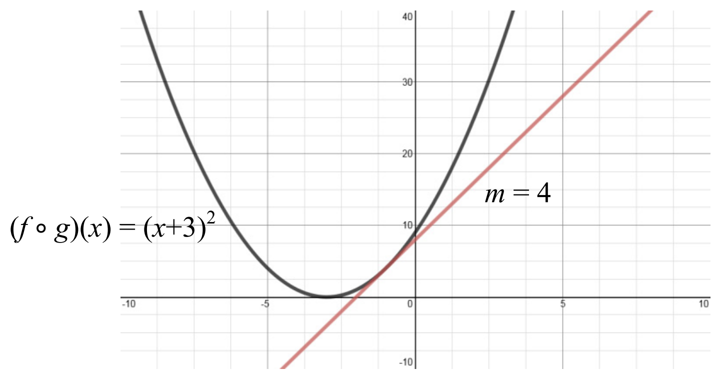



Figure 6. horizontally shifted by 3
$(f \circ g)(x) = (x+3)^2$
tangent line at $x = -1$

In this third example, we take the original graph and tangent line and horizontally shift both of these 3 units to the left. In figure 6, the new line is tangent at $x = -1$ and its slope is unchanged at 4. The process is the same as in the preceding two examples:

Letting $f(x) = x^2$, $g(x) = x+3$, and $g'(x) = 1$:

chain rule: $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$

at $x = -1$: $(f \circ g)'(-1) = f'(g(-1)) \cdot g'(-1)$

$(f \circ g)'(-1) = f'(2) \cdot 1$

$4 = 4 \cdot 1$

(transformed slope) = (original slope) · (transformation factor)

### A Limitation of the Special Case

There is a possibility that students could reach an incorrect conclusion from this special case. The notation shows that the inner derivative is to be evaluated at *x*, but linear inner derivatives are *all* constant functions. So, if students evaluated the constant inner derivatives at some location besides *x*, their error might go undetected. Teachers must help students understand this detail; if not, this special case may mislead rather than inform.

### *An Even More Special Case*

When *both* the inner and outer functions are linear, students can verify the chain rule through direct computation. Letting $g(x) = ax + b$ and $f(x) = cx + d$, for a, b, c, d $\in \mathbf{R}$, $a \neq 0$, $b \neq 0$, students can calculate that $(f \circ g)(x) = f(g(x)) = acx + bc + d$, so that $(f \circ g)'(x) = ac$. Students see the evidence that the derivative of a composite is the product of the derivatives of the components. But in this case, too, the issue of where the derivatives must be evaluated should be discussed. Both the inner and outer derivatives are constant, so their inputs do not factor into their evaluation. The general situation is very different from this.

While these special linear inner and outer cases seem very self-contained, the results here can point the way toward the chain rule in the general case, also involving the product of slopes. Students know that the tangent line does a good job of approximating a differentiable function at the point of tangency (even if they haven't yet formally studied such linearizations). So, if differentiable functions can be well approximated by a linear functions, and the chain rule for linear inner and outer functions is the product of their slopes, then it seems reasonable that the chain rule for other functions also involves the product of their slopes.

### The General Case

There is a nice way to graphically represent the process of finding the derivative of a composite in the general case, and this is available on many applets on the Internet. This helps students understand and visualize where the derivatives of the component pieces must be evaluated (Nykamp). As an example, we let the inner function $g(x) = \frac{-2}{x} + 2$, the outer function $f(x) = x^2 + 2$, and $x = 2$. We wish to calculate $(f \circ g)'(2)$. Figure 7 shows the graph of the inner function $g(x)$ and its tangent line at $x = 2$ with slope = ½, along with the graph of the outer function $f(x)$ and its tangent line at $x = 1$ with slope = 2.

Those tangent line slopes are obtained by evaluating the derivatives of $g$ and $f$ at their correct locations, and the red "cobweb" line segments, in Figure 7, show the relationship between the evaluation points. Begin at $x = 2$, and travel vertically on the red line segment until it intersects the graph of $g$. This point (2,1) is the point of tangency for $g$, and this is where $g'$ must be evaluated. Next, we must find the point of tangency of $f$, which is where $f'$ must be evaluated. The output of $g$ must become this new input, and the switch can be made by following the red horizontal line from (2,1) until it intersects the the line $y = x$. Finally, follow the red vertical line until it intersects the graph of $f$ at (1,3). This is the point of tangency for $f$, and it is where $f'$ is evaluated. Using the identity function in this way allows us to identify the input for $f'$ as the output of $g$, as needed.

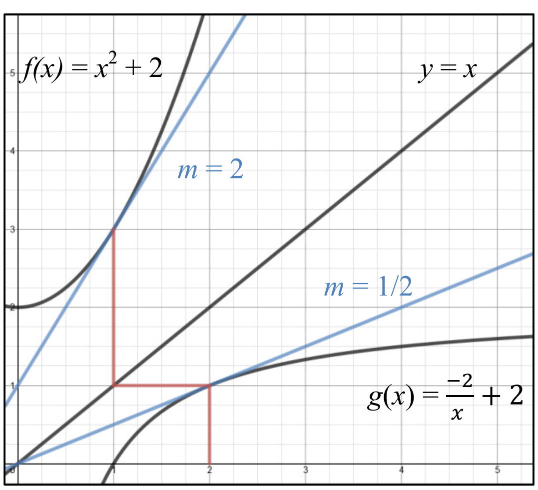


Figure 7: Two components

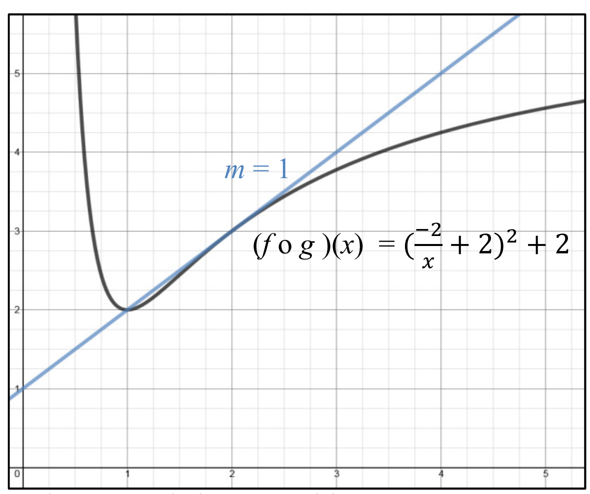


Figure 8: Their composition

In Figure 8, we can examine the graph of $(f \circ g)(x)$ and its tangent line at $x = 2$ with slope = 1. Students easily note that $m = 1$ is the product of the slopes of the inner and outer functions from Figure 7, in agreement with the chain rule.

Our earlier special cases involving linear functions help explain why the slopes in the chain rule are multiplied. This graphical technique in the general case does not help explain *why* this happens, but it does help students understand and visualize the process.

A final way to help students reinforce where the derivatives in the chain rule must be evaluted in the general case is simply to have students analyze how they compute the original composite $(f \circ g)(x) = f(g(x))$. The function $g$ is evaluated at $x$, so it is consistent that in the chain rule, $g'$ is evaluated at $x$. Additionally, the function $f$ is evaluated at $g(x)$, so it is consistent that in the chain rule $f'$ is evaluated at $g(x)$. This is helpful to students because it is easy to remember.

## The Proof of Chain Rule

We're now ready to come as close as we can to a full proof at the high school level. This common approach uses Leibniz notation:

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

$$= \lim_{\Delta x \to 0} \left[ \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x} \right]$$

$$= \lim_{\Delta x \to 0} \left[ \frac{\Delta y}{\Delta u} \right] \cdot \lim_{\Delta x \to 0} \left[ \frac{\Delta u}{\Delta x} \right] \quad \text{(since the limit of a product is the product of limits)}$$

$$= \frac{dy}{du} \cdot \frac{du}{dx}.$$

It looks like we've finally done it and it wasn't even so difficult! The good news is this is *essentially* how the actual rigorous proof works. However, students recognize the flaw in this version: $\frac{dy}{du} = \lim_{\Delta y \to 0} \left[ \frac{\Delta y}{\Delta u} \right]$, which is not the same as $\lim_{\Delta x \to 0} \left[ \frac{\Delta y}{\Delta u} \right]$. Thus, the first subsitution in the last line is invalid. Some sources and older calculus texts present an argument that addresses this flaw, claiming that since $u$ is continuous, $\Delta x \to 0$ forces $\Delta u \to 0$, so the limit on the left side above is rewritten. As a patch to the proof, however, this is also flawed, as there is nothing to prevent $\Delta u$ from being zero.

Back to the good news, though! A completely rigorous proof can be made, and it follows the spirit of this main argument. It uses a clever, but somewhat complicated, adjustment and, at this level, we're not missing much by trusting that it is correct. And after all, this flawed proof was preceeded by multiple justifications and motivations showing that the chain rule does indeed involve the product of derivatives. So, even if a full-fledged proof is somewhat beyond a class's mathematical sophistication, many lines of evidence converge on students understanding the reasons why the chain rule works.

## References


Finney, R.L., Demana, F.D., Waits, B.K., & Kennedy, D.(2003). *Calculus: Graphical, Numerical, Algebraic*. New Jersey: Prentice Hall.

Nykamp, D.Q. The idea of the chain rule. *Math Insight*. Retreived from: http://mathinsight.org/chain_rule_idea



**Joseph Gerhardinger** teaches mathematics at Notre Dame Academy in Toledo, Ohio. He stays energized by continuing to learn new mathematics. Currently, mathematical logic and Gödel's Incompleteness Theorems occupy his spare professional time.