Rethinking Rigor in Calculus: The Role of the Mean Value Theorem

Thomas W. Tucker

1. INTRODUCTION. Mathematicians have been struggling with the theoretical foundations of the calculus ever since its inception. Bishop Berkeley's attack on Newton's "ghosts of departed quantities," Euler's claim that $1 - 1 + 1 - 1 \cdots = 1/2$, Cauchy's $\varepsilon - \delta$ definition of limit, all are part of the fascinating history of this struggle (see [7]). Calculus instructors and textbooks face the same struggle, but the tack taken, although formal, is often not sensible or honest. Instead of an admission that Newton, Leibnitz, the Bernoullis, and Euler all managed quite well without any rigorous foundations, instead of the story how a rigorous calculus took mathematicians two hundred years to get right, the Mean Value Theorem is waved, like a cross in front of a vampire, to hold the difficulties at bay. The origin of the Mean Value Theorem in the structure of the real numbers is not addressed; that is much too difficult for a standard course. Maybe it is traced back to the Extreme Value Theorem, but the trail ends there. The result is that a technical existence theorem is introduced without proof and used to prove intuitively obvious statements, such as "if your speedometer reads zero, you are not going anywhere" (if $f' = 0$ on an interval, then f is constant on that interval). That's the sort of thing that gives mathematics a bad name: assuming the nonobvious to prove the obvious. And by the way, there is nothing obvious about the Mean Value Theorem without the hypothesis of continuity of the derivative. Cauchy himself was never able to prove it in that form.

I have serious reservations about the need for formal theorems and proofs in a standard calculus course. On the other hand, for those mathematicians who do feel that need, I have a suggestion for an alternative theoretical cornerstone to replace the Mean Value Theorem (MVT); I hope textbook authors adopt it. It is much easier to state, much more intuitively obvious, and much more powerful than most mathematicians realize. It is simply this:

The Increasing Function Theorem (IFT). *If $f' \geq 0$ on an interval, then f is increasing on that interval.*

Here, *increasing* means that if $c \leq d$, then $f(c) \leq f(d)$. This would usually be called nondecreasing, but that term is awkward; for example, nondecreasing and not decreasing mean different things. It seems to make more sense to use the term *strictly increasing* for the condition that if $c < d$, then $f(c) < f(d)$. A function that is increasing, but not strictly increasing, we call *weakly increasing*.

Most of the rest of this paper is concerned with the consequences of the IFT, treating it as an axiom. I will give, however, a short independent proof of the IFT, for the sake of completeness and for readers who have probably never thought of proving the IFT directly without the MVT. Of course, the IFT follows easily from the MVT. In fact, the contrapositive of the IFT is a weak form of the MVT: if $a < b$ and $f(b) < f(a)$, there is a number c, $a \leq c \leq b$, such that $f'(c) < 0$.

It is impossible to be a pioneer in territory as well-trodden as the Mean Value Theorem. Others have championed calculus without the Mean Value Theorem (see [**1**], [**4**], [**6**]). The first two sections of this paper follow Lax, Burstein, and Lax [**9**] quite closely, although unintentionally. In fact, after searching through dozens of calculus books for the Taylor remainder proof given in this paper and finally finding it in Lax-Burstein-Lax (LBL), I felt a little uncomfortable. Maybe this paper shouldn't be published and all that is needed is an announcement "Go read LBL." Then I read Grabiner [**7**] and found that the Taylor remainder proof given here and in LBL is actually Lagrange's original proof. I was surprised that such a simple, direct proof could have been covered over by years of second-growth jungle.

Moreover, the idea of Lagrange's proof keeps being rediscovered for special cases like $\sin x$ or $\cos x$. For example, the *Monthly* published such an article recently [**2**], which then generated a subsequent Editor's Note [**2**] citing calculus textbooks and *Monthly* articles where the idea of [**2**] had already been presented. None of these references noted that the same idea works for all functions; LBL is still the only book that does that, to my knowledge. And hardly anyone seems to know the idea is really Lagrange's! Under these circumstances, it appears that some dissemination is badly needed to clear up a memory lapse of generations of mathematicians. It also appears that previous calls ([**4**], [**6**]) to downplay the Mean Value Theorem have fallen on deaf ears. Perhaps the recent debates about calculus instruction have unplugged some ears and it is time to try the call again.

2. A PROOF OF THE INCREASING FUNCTION THEOREM. There is a reasonably elementary proof of the IFT that depends only on the nested interval property of the reals: if $a_n \le a_{n+1} \le b_{n+1} \le b_n$ for all $n \ge 1$ and $\lim_{n\to\infty}(b_n - a_n) = 0$, then there is a number c such that $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = c$. The proof of the IFT given here does not require the continuity of f' and is so self-contained that it probably could be given in a standard calculus course. Although I generated this proof in response to some remarks of Peter Lax, I should have known the proof is too natural to be original. In revising this paper, I discovered Richmond's article [**10**], which contains essentially the same proof, and as I already knew, Ampère and Cauchy used the key observation in their own proofs.

Proof of the IFT. The proof depends on the following simple

Observation. Given a function f, define slope(a, b) to be the usual quotient $(f(b) - f(a))/(b - a)$. If slope$(a, b) = m$ and c is between a and b, then one of slope(a, c) and slope(c, b) is greater than or equal to m and one is less than or equal to m. For a proof, draw the obvious picture.

Suppose now that $f'(x) \ge 0$ on $[a, b]$ and that f is not increasing; that is, for some a_1, b_1 with $a \le a_1 < b_1 \le b$, we have $f(a_1) > f(b_1)$. Let $m =$ slope(a_1, b_1). Note that $m < 0$. By repeated bisection and our observation, we can find a nested sequence of intervals $[a_n, b_n]$ with slope$(a_n, b_n) \le m$ and $\lim_{n\to\infty}(b_n - a_n) = 0$. Let $c = \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$ (the possibility $c = a$ or $c = b$ causes no difficulty). Since $f'(c) \ge 0$ and $m < 0$, for all x sufficiently near c, slope$(x, c) > m$. Thus for all large enough n, slope$(a_n, c) > m$ and slope$(c, b_n) > m$, which contradicts our observation and the fact that, by construction, slope$(a_n, b_n) \le m$. If $a_n = c$ or $b_n = c$, the contradiction is immediate. ∎

As we have observed, the contrapositive of the IFT is an existence statement that if f is not increasing on the interval $[a, b]$, there exists a number c between a

and b where $f'(c) < 0$. The preceding proof is constructive, in that once one finds $a_1 < b_1$ with $f(a_1) > f(b_1)$, the bisection procedure effectively computes a number c such that $f'(c) < 0$.

3. IMMEDIATE CONSEQUENCES OF THE IFT.

We first consider some consequences and variations of the IFT.

Theorem 1. *The following statements are consequences of the IFT. Assume f is differentiable on $[a, b]$ and $a < b$.*

a) *If $f'(x) \leq 0$ on the interval $[a, b]$, then f is decreasing on the interval $[a, b]$.*
b) *If $f'(x) = 0$ on the interval $[a, b]$, then f is constant on the interval $[a, b]$.*
c) *If $f'(x) > 0$ on the interval $[a, b]$, then f is strictly increasing on the interval $[a, b]$.*
d) *If $f'(x) \leq g'(x)$ on the interval $[a, b]$, then $f(x) - f(a) \leq g(x) - g(a)$ for all x in $[a, b]$.*
e) *If $m \leq f'(x) \leq M$ on the interval $[a, b]$, then $m(x - a) \leq f(x) - f(a) \leq M(x - a)$ for all x in $[a, b]$.*

Proof:

(a) Multiplication by -1 reverses inequalities and interchanges "increasing" and "decreasing".
(b) By the IFT and (a), it follows that f is both (weakly) increasing and (weakly) decreasing on $[a, b]$. That means f is constant.
(c) By the IFT, f is increasing. Suppose that $a \leq c < d \leq b$ and $f(c) = f(d)$. Since f is increasing on $[c, d]$ we must have $f(x) = f(c) = f(d)$ on $[c, d]$. That is, f is constant on $[c, d]$. Therefore $f'(x) = 0$ on $[c, d]$, contradicting $f'(x) > 0$ on $[a, b]$.
(d) Apply the IFT to $h(x) = g(x) - f(x)$ to conclude $g(a) - f(a) \leq g(x) - f(x)$.
(e) Apply (d) to $f(x)$ and Mx to get the right inequality and to mx and $f(x)$ to get the left inequality.

Theorem 1c could be called the Strictly Increasing Function Theorem (SIFT). Lax-Burstein-Lax [**9**] calls it the Criterion for Montonicity. There the IFT is derived directly from the SIFT by looking at $f(x) + mx = g(x)$, for all positive slopes m. If $f'(x) \geq 0$, then $g'(x) > 0$, so by the SIFT g is strictly increasing. Thus if $x > a$, then $f(a) + ma < f(x) + mx$. Since this inequality holds for all $m > 0$, it follows that $f(a) \leq f(x)$, that is, f is increasing. I feel, however, that this proof is a little tricky. Although the idea of perturbing a function is important throughout analysis, it comes out of the blue for a first-year calculus student. I prefer the IFT over the SIFT as a theoretical cornerstone. First, our proof that the IFT implies the SIFT is easier and more natural than a proof that the SIFT implies the IFT. More importantly, Theorem 1c, which could be called the Constant Function Theorem, follows immediately from the IFT; the only way the SIFT can get this fundamental result is via the IFT. By the way, I view the Constant Function Theorem as even more basic than the IFT. It would be nice to use it as our theoretical cornerstone, but I know of no way to use it to get the IFT.

Theorem 1d is called the Racetrack Principle by Jerry Uhl: if one car goes faster than another, it travels farther during any time interval. It is used as a theoretical cornerstone in the text [**5**].

Theorem 1e is perhaps the most important, especially from a historical viewpoint. If the inequalities are rewritten:

$$m \le \frac{f(x) - f(a)}{x - a} \le M$$

we have the Mean Value Inequality. The Mean Value Theorem follows immediately if we know that f' is continuous and that the Intermediate Value Theorem holds. That is exactly what Cauchy did [**7**]: he proved the Mean Value Inequality and assumed the continuity of f' and the Intermediate Value Theorem. His assumption of continuity should not be surprising since his proof of the Mean Value Inequality also assumes that the difference quotient $(f(x+h) - f(x))/h$ approaches $f'(x)$ uniformly as h approaches 0. Peter Lax has argued that, for the theoretical foundations of an introductory calculus course, one should always avoid pathology and assume uniform continuity and uniform convergence, just as Cauchy did. It is interesting to note that before Cauchy, Ampère [**7**] saw the importance of the Mean Value Inequality and even used it as the defining property of the derivative. One could argue in a similar vein that the Mean Value Theorem should be the defining property of the derivative; Andrew Gleason has told me that a calculus textbook by Donald Richmond around 1960 did exactly that, but I have been unable to find the book.

Finally, I should comment on the hypothesis of differentiability at the endpoints, both in the IFT and in Theorem 1. All one need assume is continuity at the endpoints, just as in the MVT. Simply observe in the proof of the IFT that the initial points a_1 and b_1 can be chosen so that $a < a_1 < b_1 < b$, since if $f(a) > f(b)$ then by continuity $f(a_1) > f(b)$ for $a_1 > a$ near enough a, and $f(a_1) > f(b_1)$ for $b_1 < b$ near enough b.

4. ERROR BOUNDS AND ERROR BEHAVIOR FOR TAYLOR POLYNOMIALS.

If Theorem 1e is rewritten

$$f(a) + m(x-a) \le f(x) \le f(a) + M(x-a),$$

we see a glimmering of an error bound for Taylor polynomials. The proof we are about to give is almost too transparent and simple to believe: just antidifferentiate repeatedly the inequality $f^{(n+1)}(x) \le M$. Not only does the proof give the Lagrange form of the error bound, it also creates the Taylor polynomial itself. Moreover, as we have observed, it is Lagrange's original proof and can be found in LBL [**9**]. It is also the proof I wrote for the textbook of the Calculus Consortium Based at Harvard [**8**]. On the other hand, I have so far been unable to find it anywhere else. All the other proofs I know involve applications of Rolle's Theorem to rather elaborate auxiliary functions or repeated integration by parts or clever tricks with varying parameters. None are natural and none are likely to be discovered or appreciated by an average calculus student.

Theorem 2. (Taylor Error Bound). *Suppose that $m \le f^{(n+1)}(x) \le M$ on the interval $[a, b]$, where $f^{(i)}$ denotes the ith derivative of f. Then on $[a, b]$*

$$m\frac{(x-a)^{n+1}}{(n+1)!} \le f(x) - T_n(x) \le M\frac{(x-a)^{n+1}}{(n+1)!},$$

where $T_n(x)$ is the degree n Taylor polynomial for f centered at $x = a$.

Proof: To get the upper bound, we apply Theorem 1d (the Racetrack Principle) to $f^{(n)}(x)$ and Mx (since $f^{n+1} \leq M$), which gives

$$f^{(n)}(x) - f^{(n)}(a) \leq M(x-a).$$

Applying the Racetrack Principle again, we get

$$f^{(n-1)}(x) - f^{(n-1)}(a) - f^{(n)}(a)(x-a) \leq M\frac{(x-a)^2}{2},$$

and again

$$f^{(n-2)}(x) - f^{(n-2)}(a) - f^{(n-1)}(a)(x-a) - f^{(n)}(a)\frac{(x-a)^2}{2} \leq M\frac{(x-a)^3}{3!}.$$

Applying the Racetrack Principle a total of $n+1$ times gives the upper bound. The lower bound is obtained the same way. ∎

Theorem 2 gives error bounds only for $x \geq a$. To get similar bounds for $x \leq a$, we observe that if f is increasing and $x \leq a$, then $f(x) \leq f(a)$, rather than $f(a) \leq f(x)$. Thus for $x \leq a$, each application of Theorem 1d reverses the inequalities, but since Theorem 2 sandwiches the error for $x \geq a$, reversing inequalities will simply sandwich the error again for $x \leq a$ (although which bound is the upper one depends on whether n is odd or even). The usual two-sided error bound involving absolute values then follows immediately.

It is possible for students to discover Theorem 2 for themselves. Consider the following problem. A particle is traveling along the x-axis with position $x = f(t)$ and suppose the initial position, velocity, and acceleration are all 0. If $f'''(t) \leq 5$ for $t \geq 0$, find an upper bound on the position at time $t = 2$. Since students are well-trained to antidifferentiate acceleration to get velocity and velocity to get position, it is not unnatural to see them argue as follows:

$$\begin{aligned} f'''(t) &\leq 5 \\ a = f''(t) &\leq 5t + c_1, \quad \text{and here } c_1 = 0 \text{ since } f''(0) = 0 \\ v = f'(t) &\leq 5\frac{t^2}{2} + c_2, \quad \text{and here } c_2 = 0 \text{ since } f'(0) = 0 \\ s = f(t) &\leq 5\frac{t^3}{6} + c_3, \quad \text{and here } c_3 = 0 \text{ since } f(0) = 0. \end{aligned}$$

Thus, we get $f(2) \leq 5 \cdot 2^3/6 = 20/3$. This is a legitimate argument as long as one can justify antidifferentiating inequalities in the same way as equalities. That is exactly the point of the Racetrack Principle!

Acceleration and velocity are not a bad way of introducing Taylor series. The usual formula students memorize from physics,

$$s = s_0 + v_0 t + \frac{1}{2}at^2,$$

is precisely the degree 2 Taylor polynomial for $s(t)$ when the constant acceleration a is interpreted as the acceleration at time 0. This fact seems worth exploiting, but I don't know any textbook that makes the connection.

Taylor's theorem is usually presented as a method of bounding the error in approximating a function by its degree n Taylor polynomial. This viewpoint is particularly appropriate in studying the error for fixed x as $n \to \infty$, as in the proof of the convergence for all values of x for the Taylor series for e^x or $\sin x$.

Nevertheless, I believe that this viewpoint is overemphasized and that the true power of Taylor series is in explaining error (or convergence) behavior for fixed n as $x \to a$. Why is Simpson's Rule so much better than the Trapezoid Rule? What makes the approximation $\sin x \approx x$ so good? For numerical behavior, the important thing to know is the order of convergence for fixed n under normal circumstances and what situations might affect that order of convergence. The real point of Taylor's theorem is that the error is order $n + 1$ in $(x - a)$ with a constant depending on the $(n + 1)^{\text{st}}$ derivative.

To be more precise, we say $E(h)$ is asymptotic to Ch^n, denoted $E(h) \sim Ch^n$, if $\lim_{h \to 0} E(h)/h^n = C$. Also, we say $E(h)$ is order n with bound M if $\limsup |E(h)/h^n\}| \leq M$. Then Taylor's theorem can be viewed this way:

Corollary. *Let $E(h)$ be the error $f(x) - T_n(x)$ where $T_n(x)$ is the nth degree Taylor polynomial for f at $x = a$ and where $h = x - a$. If $f^{(n+1)}$ is continuous at $x = a$, then $E(h) \sim f^{(n+1)}(a)\, h^{n+1}/(n+1)!$. If $|f^{(n+1)}(x)| \leq M$ in a neighborhood of $x = a$, then $E(h)$ is order $n + 1$ with bound $M/(n+1)!$.*

5. ERROR BEHAVIOR FOR NUMERICAL INTEGRATION. Another application of the Mean Value Theorem is to explain the error behavior for various common numerical integration rules: Left Rule, Right Rule, Trapezoid Rule, Midpoint Rule, Simpson's Rule. This behavior is best described using Taylor series in Δx for the error. Numerical analysis texts sometimes do this, but calculus texts don't. Since this approach is not so well-known, I'll give a version.

The idea is to concentrate on one panel of the subdivided area. Without loss of generality, we can assume the panel is centered at the origin. Thus we wish to compute

$$I(h) = \int_{-h}^{h} f(x)\, dx, \quad \text{where } h = \Delta x/2.$$

The estimate for this single panel by the left-rectangle rule is

$$I(h) \approxeq L(h) = 2h(f(-h)).$$

The other estimates are given by

$$\begin{aligned} \text{Left: } L(h) &= 2hf(-h) \\ \text{Right: } R(h) &= 2hf(h) \\ \text{Midpoint: } M(h) &= 2hf(0) \\ \text{Trapezoid: } T(h) &= (L(h) + R(h))/2 \\ \text{Simpson: } S(h) &= (2M(h) + T(h))/3 \end{aligned}$$

The formula relating Simpson's Rule to the midpoint and trapezoidal rules is not as well known as it should be. Students can be led to guess the weighted mean as a better estimate, if they spend a little time looking at the error behavior of the midpoint and trapezoidal rules.

We want to compute the Taylor series centered at $a = 0$ for all these functions. For the rules, this is simply a matter of replacing $f(h)$ or $f(-h)$ by the Taylor series for f centered at $a = 0$. For $I(h)$, we observe that by the Fundamental Theorem of Calculus, $I'(h) = f(h) + f(-h)$. Thus $I''(h) = f'(h) - f'(-h)$, $I'''(h) = f''(h) + f''(-h)$, etc.

The Taylor series for $I(h)$ is therefore

$$I(h) = 2f(0)h + 2f''(0)\frac{h^3}{3!} + 2f''''(0)\frac{h^5}{5!} + \cdots.$$

The series for the rules are

$$L(h) = 2h\left[f(0) + f'(0)(-h) + f''(0)\frac{(-h)^2}{2!} + \cdots\right]$$

$$R(h) = 2h\left[f(0) + f'(0)h + f''(0)\frac{h^2}{2} + \cdots\right]$$

$$M(h) = 2h[f(0)]$$

$$T(h) = 2h\left[f(0) + f''(0)\frac{h^2}{2} + f''''(0)\frac{h^4}{4!} + \cdots\right]$$

$$S(h) = 2h\left[f(0) + f''(0)\frac{h^2}{6} + f''''(0)\frac{h^4}{3\cdot 4!} + \cdots\right].$$

The error behavior for each rule is obtained by subtracting the Taylor series for $I(h)$ from the Taylor series for the rule and looking for the first term that doesn't cancel. The errors behave asymptotically as follows:

$$\text{Left Error} \sim -2f'(0)h^2$$

$$\text{Right Error} \sim 2f'(0)h^2$$

$$\text{Midpoint Error} \sim -2f''(0)\frac{h^3}{3!}$$

$$\text{Trapezoid Error} \sim 2f''(0)\left(\frac{1}{2} - \frac{1}{6}\right)h^3 = 2f''(0)\frac{h^3}{3}$$

$$\text{Simpson Error} \sim 2f''''(0)\left(\frac{1}{3\cdot 4!} - \frac{1}{5!}\right)h^5 = 2f''''(0)\frac{h^5}{180}.$$

The error behavior of these rules for the entire interval is obtained by multiplying by the number n of subdivisions and replacing h by $\Delta x/2$ where $\Delta x = (b-a)/n$, except for Simpson's rule where $h = \Delta x$. We have to replace $f^{(k)}(0)$ by a bound M_k on $|f^{(k)}|$ for the entire interval. Using $2nh = (b-a)$, we find the absolute value of the errors have the following behavior in terms of Δx:

Left: order 1 with bound $(b-a)(1/2)M_1$

Right: order 1 with bound $(b-a)(1/2)M_1$

Midpoint: order 2 with bound $(b-a)(1/24)M_2$

Trapezoid: order 2 with bound $(b-a)(1/12)M_2$

Simpson: order 4 with bound $(b-a)(1/180)M_4$.

The typical textbook problem on numerical integration is to find the value of n that guarantees the error is within a specified tolerance. In practice, one simply keeps doubling n until the desired number of digits seems to have stabilized. Thus, error behavior, rather than error bounds, may be what we really are interested in.

For example, it is useful to know that increasing n by a factor of 10 for the Left or Right rule, decreases the error by a factor of $1/10$, that is, it gives one more significant digit. Thus if it takes 1 second for a graphing calculator to compute an integral accurate to 2 digits using the Left or Right Rule, it will take 10^{10} seconds to get 12 digits of accuracy (that's 3169 years and, as my students have observed, a lot of batteries). By contrast, Simpson's Rule gets 4 extra digits for 10 times the work, and the same integral can be computed to 12 digits of accuracy in a minute or two on the same calculator (Simpson's Rule probably would get a headstart of 4 or 5 digits in the first second).

The dependence of error behavior on the higher derivatives of the integrand is also important, because it is a warning to look out for integrals whose integrand has an unbounded derivative on the interval of integration. For example, even using Simpson's Rule on $\int_0^1 \sqrt{1-x^2}\,dx$ to get an approximation for $\pi/4$ is painfully slow going. Indeed, the order of convergence is $3/2$ rather than 4.

Taylor series can be used in the same way to analyze the error behavior for numerical differentiation approximations:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}$$

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

$$f''(x) \approx \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

For example, students are often curious why some graphing calculators use the second of these approximations as a numerical derivative rather than the more familiar first approximation. Taylor series give the answer immediately: the second error for the approximation is order 2 while the first is order 1. The dependence of the error of each approximation on higher derivatives of f also has interesting effects. Try plotting the error near $x = 0$ with $h = .01$ for the second approximation to f', when f is the innocuous-looking function $f(x) = x^{8/3}$.

6. THE FUNDAMENTAL THEOREMS OF CALCULUS. The proof given in [**8**] for the Taylor error bound appeals to the Fundamental Theorem of Calculus to turn the inequality $f^{n+1}(x) \leq M$ into the inequality $f^{(n)}(x) - f^{(n)}(a) \leq M(x - a)$. I suspect this is the natural inclination of most mathematicians, and it shows how much under-appreciated the IFT is. No definite integrals are needed; the IFT itself is a disguised form of integration. The subtle connection between the IFT and the Fundamental Theorem of Calculus is worth discussing.

There are of course two main versions of the Fundamental Theorem of Calculus. There are also variations on what restrictions are placed on the integrand f. I will assume f is continuous. The theorems then are

First Fundamental Theorem of Calculus (FTC I). *If f is continuous on the interval $[a, b]$ and $F(x) = \int_a^x f(t)\,dt$ for x in $[a, b]$, then $F'(x) = f(x)$.*

Second Fundamental Theorem of Calculus (FTC II). *If f is continuous and $F(x) = f(x)$, then $\int_a^b f(t)\,dt = F(b) - F(a)$.*

The First Fundamental Theorem is not directly related to the IFT. The hard part of the proof is showing that continuous functions are Riemann integrable. The rest is a straightforward consequence of the integral version of the Mean Value Inequality:

$$m(b-a) \leq \int_a^b f(x)\,dx \leq M(b-a),$$

where $m \leq f(x) \leq M$ on the interval $[a, b]$. Note that unlike the Mean Value Inequality for derivatives, this inequality follows easily from the definition of the Riemann integral, so easily that it is not uncommon to view the inequality as a defining property of the definite integral (the corresponding view for the Mean Value Inequality for derivatives, Ampère notwithstanding, is much less common).

On the other hand, the Second Fundamental Theorem is closely connected to the IFT. The IFT for continuously differentiable functions follows directly from the FTC II and the fact that the integral of a nonnegative function is nonnegative. In fact, that is the way the IFT is proved in [**8**]. There, the FTC II, as embodied in the relation between velocity and change in position, is taken as the intuitively clear, theoretical cornerstone, and the IFT is derived from it. I suspect, however, that most students see the IFT as more "obvious" than the FTC II.

Conversely, the IFT implies the FTC II by the method used in many calculus books: simply invoke the FTC I with $x = b$ and observe that, by the IFT (the Constant Function Theorem, Theorem 1b), two antiderivatives of f differ by a constant.

The assumption of continuity in the FTC I is necessary. The assumption of continuity in the FTC II is another matter. Of course, if $F' = f$ is not continuous, the integral might not exist. For example, if $F(0) = 0$ and $F(x) = x^2 \sin(1/x^2)$ for $x \neq 0$, then F' exists everywhere but is not even Lebesgue integrable on $[0, 1]$. Suppose, however, that we assume only that $\int_a^b f(t)\,dt$ exists. Then the familiar argument using the Mean Value Theorem still works. Just represent $F(b) - F(a)$ as a telescoping sum and use the MVT on each term of the sum to turn it into a Riemann sum for $\int_a^b f(t)\,dt$. Here the IFT does not work. Just as the MVT follows from the IFT only under the assumption of continuity of the derivative, the FTC II follows from the IFT only under the assumption of continuity of the integrand.

7. CONCLUSION. Many calculus textbooks have sections where the author is writing on automatic pilot, just putting in material demanded by users. These sections have the same dreary examples; little is new, or thought over fresh from the start. This shouldn't be surprising, since writing a calculus textbook is a significant project and one can't devote the same enthusiasm and energy to all parts of the project. I have always felt that the theoretical sections of standard calculus textbooks are most prone to such a pedestrian treatment. Moreover, calculus instruction does not place much emphasis on those theoretical sections, at least when it comes to testing. For example, a study of the compendium of final exams in [**11**] reveals only one question (out of more than 300 on 23 exams) involving the Mean Value Theorem, and that one asked for the value of c satisfying the conclusion of the Mean Value Theorem for a quadratic function. When both textbooks and instruction appear to be just going through the motions with theory, it surprises me that some critics of new textbooks like [**8**] bemoan the absence of the Mean Value Theorem or a $\epsilon - \delta$ definition of limit.

I sympathize with yearnings for an occasional foray into the theoretical structure of the calculus. I just ask that it be thoughtful and sensible. Use intuitive definitions. If a theorem is to be used without proof, like the Mean Value Theorem, keep it as simple and as "obvious" as possible. Don't use tricky proofs or deus-ex-machina auxiliary functions. Don't prove things in more generality than necessary; even analysts don't usually deal with the discontinuous derivatives allowed by the Mean Value Theorem.

In this paper, I have tried to give a sensible approach to the Mean Value Theorem and its usual applications to monotonicity, Taylor error bounds, quadrature error bounds, and the Fundamental Theorems of Calculus. One standard application of the MVT I have not considered is l'Hopital's Rule; for a non-MVT approach, see [3]. LBL [9] has some other applications to concavity and the second derivative test for extrema.

In recent years, calculus content and pedagogy have been rethought completely. People have found that there is nothing sacred about related rates and the lecture method. It is time as well to rethink the theory taught in standard calculus classes. There is nothing sacred about the Mean Value Theorem.

ACKNOWLEDGMENTS. I wish to thank Andy Gleason, Peter Lax, and Jerry Uhl for numerous suggestions and corrections for this article. In particular, the proof given for the IFT was instigated by a bisection proof Lax showed me for the SIFT. He also showed me applications to l'Hopital's Rule, the Corrected Midpoint Rule for quadrature, and the definition of volumes and arclengths using antiderivatives rather than definite integrals; all of this I hope he puts into print. I am indebted to Gleason, whose meticulous reading caught a number of egregious errors and whose comments cleared up my muddy thinking at numerous points.

REFERENCES

1. L. Bers, On avoiding the mean value theorem, *Amer. Math. Monthly* 74 (1967), 583.
2. D. Bo, A simple derivation of the Maclaurin series for sine and cosine, *Amer. Math. Monthly* 97 (1990), 836. Editor's Note in the *Monthly* 98 (1991), 364.
3. R. P. Boas, Lhospital's rule without mean value theorems, *Amer. Math. Monthly* 76 (1969), 1051–1053.
4. R. P. Boas, Who needs these mean-value theorems anyway?, *Two-Year College Math J.* 12 (1981), 178–181.
5. W. Davis, H. Porta, J. Uhl, *Calculus & Mathematica*: *Derivatives*: *Measuring Growth*, Addison-Wesley, 1994.
6. J. Dieudonne, *Foundations of Modern Analysis*, Academic Press, New York, 1960.
7. J. V. Grabiner, *The Origins of Cauchy's Rigorous Calculus*, MIT Press, Cambridge, 1981.
8. D. Hughes-Hallett, A. M. Gleason, et al., *Calculus*, John Wiley & Sons, New York, 1994.
9. P. Lax, S. Burstein, and A. Lax, *Calculus with Applications and Computing*, *Volume* 1, Springer-Verlag, New York, 1984.
10. D. E. Richmond, An elementary proof of a theorem of calculus, *Amer. Math. Monthly* 92 (1985), 589–590.
11. L. A. Steen, editor, *Calculus for a New Century*: *A Pump*, *Not a Filter*, MAA Notes 8, Mathematical Association of America, Washington, DC, 1988.

Department of Mathematics
Colgate University
Hamilton, NY 13346
ttucker@center.colgate.edu