Rethinking the Elementary Real Analysis Course

Brian S. Thomson

1. INTRODUCTION.

> Finally, the reader will probably observe the conspicuous absence of the time-honored topic in calculus courses, the "Riemann integral." It may well be suspected that, had it not been for its prestigious name, this would have been dropped long ago, for (with due reverence to Riemann's genius) it is certainly quite clear to any working mathematician that nowadays such a "theory" has at best the importance of a mildly interesting exercise in the general theory of measure and integration. Only the stubborn conservatism of academic tradition could freeze it into a regular part of the curriculum, long after it had outlived its historical importance. Of course, it is perfectly feasible to limit the integration process to a category of functions which is large enough for all purposes of elementary analysis, but close enough to the continuous functions to dispense with any consideration drawn from measure theory; this is what we have done by defining only the integral of regulated functions. When one needs a more powerful tool there is no point in stopping halfway, and the general theory of (Lebesgue) integration is the only sensible answer.
>
> Jean Dieudonné, *Foundations of Modern Analysis* (1960)

This well-known quotation from a formidable mathematician with an equally formidable personality should, by now, have rung the death knell for the Riemann integral. North American calculus courses, naturally, have not responded, but one could have expected no further elementary analysis texts to appear with a "serious" treatment of an inadequate integral. The sentiment, however, seems to be that, since measure theory would be inappropriate in a first analysis course, the only alternative is to develop the Riemann integral. Certainly, the student will set it aside at some later date, like training wheels on a bicycle, but how else can any integration theory be done?

Even in 1960, when Dieudonné launched this attack, there was an alternative available. The full range of classical integration theories on the real line (Newton, Riemann, Lebesgue, Denjoy-Perron) had found a simple formal expression that could easily be developed in a suitable manner for a first course. The Henstock-Kurzweil integral, as it came to be known (see [**5**] and [**7**]), is not just easier to present (at least at the early formal stages) than the Lebesgue integral. It is easier to present than the Riemann integral. It is also, arguably, the *correct* integral for the calculus program, but that might be harder for the reader to accept.

The time may be right again for a return to these ideas. The late Bob Bartle joined the ranks of the supporters of the alternative integration theory by publishing a note in this MONTHLY [**1**]. As this received the Mathematical Association of America's Lester R. Ford Award, we can presume that there is now greater sentiment for trying out such a program.

In this article we address some of the concerns that a designer of an elementary real analysis course should confront when choosing to throw off the Riemann integral. Our thesis is not at all that the Henstock-Kurzweil integral should be included. Instead we promote a simpler viewpoint: that the use of Cousin covering arguments could be made the centerpiece of the course. If that is accepted, the right integration theory is inescapable and emerges naturally and simply.

2. AN ANALYSIS EXERCISE IN SEARCH OF A METHOD. The typical calculus text of our times avoids, or consigns to a remote appendix, any use of compactness arguments. In particular then, while much use is made of the facts that continuous functions have maxima and minima on compact intervals and possess the Darboux property, such facts are beyond proof. A first analysis course is invariably called upon to make these compactness arguments. But which form should take center stage?

We illustrate the question with an exercise appropriate to the level.

1. A function $f : \mathbb{R} \to \mathbb{R}$ is *constant on a set* E if $f(x) = f(y)$ for all x and y in E.
2. A function $f : \mathbb{R} \to \mathbb{R}$ is *locally constant at a point* x if there is a $\delta > 0$ so that $f(x) = f(y)$ for every y with $|x - y| < \delta$.

Exercise 1. Show that a function that is locally constant at each point of an interval $[a, b]$ is constant on $[a, b]$.

A calculus student might "solve" this by spotting that such a function would have to have a zero derivative everywhere. Thus, by a familiar calculus principle, the function is constant. But the proof of that principle requires a compactness argument anyway, and the exercise should be attempted from first principles.

The following arguments are all equivalent since any one of them can be used to prove the others. In spite of the equivalence, some have advantages over the others.

Sup/Inf Argument. Define the set C as follows:

$$C = \{x \in [a, b] : \ f(t) = f(a) \text{ whenever } a \leq t \leq x\}.$$

Verify that C is nonempty and bounded. The sup/inf principle demands that C have a supremum (a least upper bound). The existence of the supremum and the fact that f is locally constant at that point will solve the exercise.

This principle is usually taken as an axiom. It is, in any case, an inescapable tool in elementary analysis.

Dedekind Cuts. If f is not constant then there are points x and y in $[a, b]$ $(x < y)$ with $f(x) \neq f(y)$. Define the sets A and B by

$$A = \{t : \ t \leq x \text{ or } x < t \text{ and } f \text{ is constant on } [x, t]\ \}$$

$$B = \{t : \ x < t \text{ and } f(z) \neq f(x) \text{ for some } x < z \leq t\ \}$$

Verify that A and B are nonempty and exhaust the real line. Dedekind's principle asserts that there is a cut number, a number separating the two sets. That number must be in the interval $[x, y]$. The fact that f is locally constant at that point will show that it cannot belong to either set A or B. This contradiction solves the exercise.

Nested Interval Argument. Suppose that f is not constant on $[a, b]$. Subdividing that interval in equal subintervals, we produce at least one subinterval $[a_1, b_1]$ of $[a, b]$ with half the length on which, again, f is not constant. We continue inductively, producing a nested sequence of intervals whose lengths shrink to zero. The nested interval principle ensures that there is a unique number that belongs to all of the intervals. The fact that f is locally constant at that point produces a contradiction for the intervals that have been constructed.

Sequential Compactness Argument. If f is not constant on an interval $[a, b]$, then there are points x_1 and y_1 from that interval with $f(x_1) \neq f(y_1)$ and $|x_1 - y_1| \leq 1$. Argue so as to produce points x_n and y_n from $[a, b]$ with $f(x_n) \neq f(y_n)$ and $|x_n - y_n| \leq 1/n$. The Bolzano-Weierstrass property of sequences in compact sets allows us to extract a convergent subsequence from $\{x_n\}$, and hence a point which is a limit of subsequences from both $\{x_n\}$ and $\{y_n\}$. The fact that f is locally constant at that point leads to a contradiction to the way in which the sequences have been constructed.

Heine-Borel Argument. Let $\mathcal{G}$ be the collection of all open intervals on which f is constant. By the Heine-Borel principle there must a finite collection of intervals from $\mathcal{G}$ that covers $[a, b]$, so f would have to be constant on $[a, b]$.

This is a first-class compactness argument whose greatest merit is that it is exactly the technique used in most advanced settings. On the other hand, our goal is not to select the technique used by professionals, but to select the one that is most useful and simple for an elementary analysis course.

Lebesgue Chains. We note that for every point x of $[a, b)$ there is an interval $[x, x + h]$ on which f is constant. A chain of such intervals (see [**10**]) joining a to b will verify that f is constant.

Start off with

$$[a, a_1], [a_1, a_2], [a_2, a_3], \ldots$$

always selecting intervals on which f is constant. Eventually one reaches b in a countable numbers of steps. Well, more correctly, if $a_\omega = \lim_n a_n$ is less than b then carry on with $[a_\omega, a_{\omega+1}], \ldots$ in a transfinite sequence. Transfinite induction, using local constancy at each limit ordinal, completes the proof.

Perhaps this might be a bit too mysterious for a first course in analysis! In fact, though, one of the covering arguments we ultimately use is closely related to this, but has the decided advantage of replacing a transfinite sequence of intervals with a finite one.

Covering Argument. We think about the nature of locally constant functions by considering pairs (I, x), where I is a closed interval containing the point x. Collections of pairs are called *relations* and, because we use only pairs (I, x) with I an interval containing x, these are commonly called *covering relations*.

The pairs that describe the situation of the exercise are these:

> All pairs (I, x), where I is a closed subinterval of $[a, b]$, x is a point of I, and f is constant throughout the whole interval I.

Visualize the collection β of all such pairs: in your imagination you see many little pieces describing how the function is constant close to each point. The collection β is a covering relation that describes the geometry of our problem. The covering relation β has the following simple local structure: if x is an arbitrary point of $[a, b]$, there must be a $\delta > 0$ so that a pair (I, x) belongs to β whenever x belongs to I and the length of I is smaller than δ.

The resolution of the problem depends on putting the pieces together. The Cousin covering lemma (Lemma 1) asserts there is a partition of the interval $[a, b]$ formed of a finite number of elements (I_i, x_i) from β. On each interval I_i the function f is constant. Since the intervals $\{I_i\}$ cover $[a, b]$, it follows that f is constant on the entire interval.

Selecting the Right Compactness Argument. The choice of the Bolzano-Weierstrass property or the Heine-Borel property would seem to be dictated, both by the needs of more advanced courses that the student will subsequently take and by the obvious predilection of the instructor for the more familiar methods that he uses daily. Admittedly, the covering argument producing partitions of compact intervals is very peculiar to the geometry of the real line. No such geometry would be available in metric spaces or general topological spaces.

Still, it is the very fact that the method is tailored to apply to the real line that gives it its merit. Much of the nature of the study of derivatives, integrals, and limits on the real line can be expressed by these covering notions. Many of the properties of these concepts are clarified by the two covering lemmas in the next section. We ask the reader to accept this premise for the moment in order to see where it leads.

The name of the covering lemma originates from a paper in the late nineteenth century by the Belgian mathematician Pierre Cousin. It seems entirely unreasonable to attribute the lemma to him, but so far no earlier volunteer has stepped forward. For a while it looked like his more famous French colleague, Edouard Goursat, could claim the honors for a paper published in 1900, but Cousin's paper was found and predates Goursat's by a little bit.

The lemma has a peculiar and persistent habit of rediscovery. An avid reader of twentieth century copies of this MONTHLY will find that it reappears with nearly the regularity and twice the frequency of the cicadas. A friend of mine who used it in a basic analysis course claims that the lemma is perhaps *too* successful in early courses: the students become so enamored in using the machinery of the lemma that they avoid learning any other techniques.

3. COVERING RELATIONS.

A *covering relation* is a family β of interval-point pairs $([a, b], c)$, where $a < b$ and c is an element of $[a, b]$. A *covering lemma* is a statement that from some covering relation β a subset β_1 can be extracted with certain desired properties.

The most famous of covering lemmas is due to Vitali and can be presented only with a good bit of measure theory, not what we have in mind for this elementary development. The simplest and most elementary of covering lemmas are those that concern covering relations arising in connection with compact sets.

Definition 1. A covering relation β is a *Cousin cover* of a compact interval $[a, b]$ if for each x in $[a, b]$ there is a $\delta > 0$ with the property that $([c, d], x)$ belongs to β whenever $x \in [c, d] \subset [a, b]$ and $d - c < \delta$.

While the Cousin covers are designed to address the situation that we saw in our elementary exercise, we require also some more general covers that require only a weak condition to hold on the right-hand side at each point provided a strong condition holds on the left. These covers first arose in a study of right-hand derivatives of continuous functions in Hagood [**3**], where they are used to give an elementary proof of the Lebesgue differentiation theorem. We shall reproduce that proof in Section 7.

Definition 2. Let K be a compact subset of the real numbers with $a = \inf K$ and $b = \sup K$. A covering relation β is a *quasi-Cousin cover* of K provided that the following conditions are met:

1. There is at least one pair $([a, d], a)$ in β with $a < d \leq b$.
2. For each x in $K \cap (a, b)$ there is a $\delta > 0$ and a d satisfying $x < d \leq b$ such that $([c, d], x)$ belongs to β whenever $x - \delta < c \leq x$.

3. There is a $\delta > 0$ so that $([c, b], b)$ belongs to β whenever $b - \delta < c < b$.

Our two main covering lemmas establish that Cousin and quasi-Cousin covering relations contain partitions. Suppose that

$$a = a_0 < a_1 < a_2 < \cdots < a_{k-1} < a_k = b$$

and that x_i is a point in $[a_{i-1}, a_i]$ for each $i = 1, 2, \ldots, k$. Then we would call the covering relation

$$\pi = \{([a_{i-1}, a_i], x_i) : i = 1, 2, \ldots, k\}$$

a *partition* of $[a, b]$. Any subcollection of a partition is called a *subpartition*. Note that for a covering relation π to be a subpartition it need only be a finite collection such that the intervals I_1 and I_2 do not overlap for distinct pairs (I_1, x_1) and (I_2, x_2) in π.

Lemma 1 (Cousin Covering Lemma). *If β is a Cousin cover of a compact interval $[a, b]$, then β contains a partition of every compact subinterval of $[a, b]$.*

Lemma 2 (Quasi-Cousin Covering Lemma). *If β is a quasi-Cousin cover of a compact set K with $a = \inf K$ and $b = \sup K$, then β contains a subpartition π such that*

$$K \subset \bigcup_{(I,x)\in\pi} I \subset [a, b].$$

In particular, if K is itself a compact interval $[a, b]$, then β must contain a partition of $[a, b]$. The most economical line of proof is to establish Lemma 2 first using a sup/inf argument and then appeal to Exercise 4 to prove Lemma 1. (The exercises are given for reference, not as assignments to the reader.)

Exercise 2. Let K be a compact set, and let β be a covering relation with the property that for each x in K there exist positive numbers s and t so that $([x', x + s], x)$ belongs to β whenever $x - t \le x' \le x$. Show that β contains a subpartition π for which

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

Exercise 3. Show that if β_1 and β_2 are both Cousin covers of $[a, b]$, then $\beta_1 \cap \beta_2$ is also a Cousin cover of $[a, b]$.

Exercise 4. Show that every Cousin cover of an interval $[a, b]$ is a quasi-Cousin cover of any subinterval $[c, d]$ of $[a, b]$.

Constructing Covering Arguments. The key advice in constructing covering arguments is to keep things simple. Make sure that a cover reflects the geometry of the problem. It is an error to construct highly technical devices with lots of ϵs and δs. The only technical step in each argument should be the actual checking of the fact that the cover satisfies the hypotheses of the lemma invoked. The construction of the covering relation and its role in solving the problem should always be transparent.

There will be a grade of "C−" if the description of the covering relation is framed in such a way to verify the fact that it is a Cousin cover. The literature abounds with clumsy formulations of these ideas. If you see mention of anything being "δ-fine" you can consider yourself in "C−" territory.

Example 1. Let $\mathcal{G}$ be a family of open sets with the property that each point of a compact set K is contained in at least one member of the family. Show that there is a finite subcollection $\mathcal{G}_0$ of $\mathcal{G}$ with the same property.

The geometry of this problem is expressed in the covering relation

$$\beta = \{(I, x) : x \in I \text{ and } I \subset G \text{ for some } G \text{ in } \mathcal{G}\}.$$

Just check that β is a quasi-Cousin cover of K. The subpartition guaranteed by Lemma 2 immediately solves the problem.

Example 2. Let E be an infinite, bounded set. Show that E must have a point of accumulation.

The geometry of this problem is captured by the covering relation

$$\beta = \{(I, x) : x \in I \text{ and } I \cap E \text{ is finite}\}.$$

Assuming that E has no point of accumulation, check that β is a Cousin (or quasi-Cousin) cover of any compact interval.

Example 3. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. Show that f is uniformly continuous on every compact subinterval of $\mathbb{R}$.

Let $\epsilon > 0$. The geometry of the problem reduces to the covering relation

$$\beta = \{(I, x) : x \in I \text{ and } \omega_f(I) < \epsilon\},$$

where $\omega_f(I) = \sup_{x,y \in I} |f(x) - f(y)|$ is the oscillation of the function f on the interval I.

Example 4. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function with $f'(x) = 0$ at all but a countable number of points. Show that f is constant.

The geometry of this problem involves two covering relations, namely,

$$\beta_1 = \{([y, z], x) : x \in [y, z] \text{ and } |f(z) - f(y)| < \epsilon(z - y)/2\}$$

and

$$\beta_2 = \{(I, x) : x = c_i \in I \text{ and } \omega_f(I) < \epsilon 2^{-i-1}\}.$$

Here $\{c_1, c_2, c_3, \dots\}$ is the set of exceptional points at which we do not know the existence of the derivative. The union $\beta_1 \cup \beta_2$ is a Cousin (or quasi-Cousin) cover of any compact interval. The promised partition can be split into two parts for easy handling.

Example 5. Let $f : \mathbb{R} \to \mathbb{R}$ be a function that satisfies the Lipschitz condition $|f(x) - f(y)| \leq M|x - y|$ for a positive real number M and all real numbers x and y. Suppose that $f'(x) = 0$ at all x excepting only a set of measure zero. Show that f is constant.

Although the statement of the exercise would seem to require some measure theory, its solution can be accomplished with very little. All we need is that for $\epsilon > 0$ there is an open set G whose Lebesgue measure is smaller than ϵ and $f'(x) = 0$ for all x not in G.

The geometry of this problem is again expressed in two covering relations:

$$\beta_1 = \{([y,z],x) : x \in [y,z] \text{ and } |f(y) - f(z)| < \epsilon(y-z)/2\}$$

and

$$\beta_2 = \{(I,x) : x \in I \text{ and } I \subset G\},$$

where G is an open set chosen so that the Lebesgue measure of G is smaller than $\epsilon/(2M)$, and where $f'(x) = 0$ if x is not in G. (For a more ambitious exercise, with nearly the same proof, advanced readers may assume instead that f is absolutely continuous.)

4. INTEGRALS. The integral of choice for our course is determined by two factors: the inversion of derivatives and the covering relation that expresses that process. The driving force behind the development of the integral on the real line has always been the problem of recovering a function from its derivative. Some presentations of Lebesgue's integral can easily obscure this, even though Lebesgue himself claimed it was his motivation. Thus one can find textbook complaints about the Riemann integral that it lacks appropriate limiting properties or doesn't integrate "enough" functions, but with no genuine indication of why more functions would be needed.

By contrast, a natural and compelling motivation is offered by a return to the familiar calculus problem of inverting a derivative. The following lemma is an entirely elementary expression of the geometry of the relation $F'(x) = f(x)$ in the language of covering relations. (We use I to denote a compact interval, $\ell(I)$ to denote its length, and $\Delta F(I)$ to denote the increment of a function F on I.)

Lemma 3. *If $F : \mathbb{R} \to \mathbb{R}$ is a differentiable function with $F'(x) = f(x)$ everywhere, then for each $\epsilon > 0$ and compact interval $[a,b]$ there is a Cousin cover β of $[a,b]$ such that*

$$\left| \sum_{(I,x)\in\pi} f(x)\ell(I) - \Delta F([a,b]) \right| < \epsilon$$

and

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\ell(I)| < \epsilon$$

for every partition π of $[a,b]$ contained in β.

The proof is trivial: merely take

$$\beta = \left\{ (I,x) : \ x \in I \text{ and } \left| \frac{\Delta F(I)}{\ell(I)} - f(x) \right| < \epsilon/(b-a) \right\}.$$

But the lemma contains a serious clue as to how one can (formally) recover $\Delta F([a,b])$ from f, as well as the fundamental characterization of the integral itself (see Henstock's criterion in Section 5).

The Integral. A function f is *integrable* (in the sense appropriate to the present discussion) on a compact interval $[a, b]$ if there is a number c so that for each $\epsilon > 0$ a Cousin cover β of $[a, b]$ can be found with the property that

$$\left| \sum_{(I,z)\in\pi} f(z)\ell(I) - c \right| < \epsilon$$

for all partitions π of $[a, b]$ contained in β. This number c is denoted, naturally, by the symbol $\int_a^b f(x)\,dx$.

This definition is just a copy of the property of derivatives expressed in Lemma 3. The definition is designed to accomodate derivatives, and so it does: it is immediate (from Lemma 3) that, if $F' = f$ everywhere, then f is integrable on any interval $[a, b]$ and

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Upper and Lower Integrals. For most presentations it is preferable to introduce upper and lower integrals and then deduce the definition just given as a theorem.

Definition 3. For a function f defined on an interval $[a, b]$ we define an upper integral by

$$\overline{\int_a^b} f(x)\,dx = \inf_{\beta} \sup_{\pi\subset\beta} \sum_{(I,z)\in\pi} f(z)\ell(I),$$

where the supremum is taken over all partitions π of $[a, b]$ contained in a fixed Cousin cover β of $[a, b]$ and the infimum extends over all such covers.

Similarly, we define a lower integral as a sup/inf. If the upper and lower integrals are identical we write the common value as $\int_a^b f(x)\,dx$. If this value is also finite, then we say f is *integrable*. It is easy to check that the two methods of defining the integral agree.

5. PROPERTIES OF THE INTEGRAL. The integral (it is advised to call it merely *the integral* rather than attach some person's name to it) can now be developed as far as is needed for the course design. Those who are of the sentiment that their students need scarcely anything more than the Riemann integral just don't have to go so deep. If the goal is to get the student as far only as the rudiments of the Lebesgue integral, then again push only that far. The view that this integral is peculiar or recondite should be suppressed. Restricted to the class of functions that you usually use this is just the same integral; the fact that more functions can be integrated simplifies many arguments and assertions, for unnecessary hypotheses can be dropped and no longer play roles in proofs. We have seen one example already: if $F'(x) = f(x)$ everywhere, then f is integrable over each compact interval $[a, b]$ and

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

The proof, as we have seen, is completely trivial. In the setting of the Riemann integral this is false without also assuming that f be integrable. That complicates the proof.

In the setting of the Lebesgue integral, again integrability must be assumed and the considerable apparatus of measure theory then used to verify the identity.

The elementary properties of the integral can be stated and proved much as in a typical analysis course. Exercise 3 plays a role in most of the easy proofs, but otherwise things look quite familiar. While it has a moderately more complicated definition than the Riemann integral, this integral has much the same structure and the simplest properties flow from that fact.

Improper Integrals? In the usual calculus course, we treat an integral such as $\int_0^1 x^{-1/2}\,dx$ by the familiar device:

$$\int_a^b f(x)\,dx = \lim_{c \searrow a} \int_c^b f(x)\,dx. \tag{1}$$

Here, if the left-hand side does not exist in the usual sense it is to be interpreted as the right-hand side, provided that exists.

It is not because the integrand is undefined at the endpoint that this is needed (as some students think): values of a function at a single point do not influence integrability in a Riemann-type theory and so any value could be assumed at a troublesome point. The calculus integral applies only to bounded functions. This is the reason that the student is obliged to announce the existence of the integral in "an improper sense" and justify that existence by several comments.

The integral that we have defined needs no such justifications. The identity in (1) always holds: the improper extension of the integral is no longer needed. The course designer should keep in mind that the Lebesgue integral does not have this property and different formulations are needed to check integrability at points of unboundedness.

Establishing this property is not difficult but could be avoided in a simple course. For most applications (certainly for the integral $\int_0^1 x^{-1/2}\,dx$) the following lemma would be preferred anyway:

Lemma 4. *If F is a continuous function on an interval $[a, b]$ for which $F'(x) = f(x)$ everywhere there with at most countably many exceptions, then f is integrable on $[a, b]$ and*

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

The proof is easy: verify that that the conclusion of Lemma 3 holds (mimic the covering argument in Example 4).

Summing Inside the Integral. We wish now to establish the formula

$$\int_a^b \left(\sum_{n=1}^{\infty} f_n(x)\right) dx = \sum_{n=1}^{\infty} \left(\int_a^b f_n(x)\,dx\right).$$

Such a formula would never be proved in a first course for the Riemann integral except under the assumption of uniform convergence. For the Lebesgue integral the proof would require some serious measure-theoretic tools. It is part of the tradition of the Henstock-Kurzweil integral that such theorems can be handled without many preliminaries; the version here, using upper and lower integrals, is particularly simple because so few hypotheses are involved.

The covering argument chops up and then puts together the fragments of Cousin covers; while somewhat detailed, this should be accessible in a first course.

Theorem 1. *Suppose that f and f_n $(n = 1, 2, \dots)$ are nonnegative functions on a compact interval $[a, b]$. If*

$$f(x) \le \sum_{n=1}^{\infty} f_n(x)$$

for each x in $[a, b]$, then

$$\overline{\int_a^b} f(x)\,dx \le \sum_{n=1}^{\infty} \left(\overline{\int_a^b} f_n(x)\,dx \right). \tag{2}$$

If

$$f(x) \ge \sum_{n=1}^{\infty} f_n(x)$$

for each x in $[a, b]$, then

$$\underline{\int_a^b} f(x)\,dx \ge \sum_{n=1}^{\infty} \left(\underline{\int_a^b} f(x)\,dx \right). \tag{3}$$

Proof. Inequality (3) is the easiest to prove: armed with Exercise 3 the reader will have no troubles. To prove (2) requires some manipulations of the covering relations and a bit of bookkeeping. Let $t < 1$ and $\epsilon > 0$. For x in $[a, b]$ let $N(x)$ signify the first integer such that

$$tf(x) \le \sum_{n=1}^{N(x)} f_n(x).$$

Choose Cousin covers β_n of $[a, b]$ $(n = 1, 2, \dots)$ so that $\beta_1 \supset \beta_2 \supset \beta_3 \dots$ and

$$\sum_{\pi} f_n(x)\ell(I) \le \overline{\int_a^b} f_n(x)\,dx + \epsilon 2^{-n}$$

whenever π is a partition of $[a, b]$ contained in β_n. (If any of the integrals here is not finite then there is nothing to prove, since the right-hand side of (2) will be infinite.)

Let

$$E_n = \{x \in [a, b] : N(x) = n\}.$$

We use these sets to split up the covering relations. Write

$$\beta_n[E_n] = \{(I, x) \in \beta_n : x \in E_n\}$$

and

$$\beta = \bigcup_{n=1}^{\infty} \beta_n[E_n].$$

It is straightforward to check that β is a Cousin cover of $[a, b]$. Consider any partition π of $[a, b]$ contained in β. Let N be the largest value of $N(x)$ for the finite collection of pairs (I, x) in π. We need to divide the partition π into a finite number of subpartitions by writing

$$\pi_j = \{(I, x) \in \pi : x \in E_j\}$$

and

$$\sigma_j = \pi_j \cup \pi_{j+1} \cup \cdots \cup \pi_N$$

for $j = 1, 2, \ldots, N$. Note that σ_j is a subcollection of β_j and that $\pi = \pi_1 \cup \pi_2 \cup \cdots \cup \pi_N$.

Keeping in mind that

$$tf(x) \leq f_1(x) + f_2(x) + \cdots + f_i(x) \ (x \in E_i),$$

we check the following computations:

$$\begin{aligned}
\sum_{\pi} tf(x)\ell(I) &= \sum_{i=1}^{N}\left(\sum_{\pi_i} tf(x)\ell(I)\right) \\
&\leq \sum_{i=1}^{N}\sum_{\pi_i}[f_1(x) + f_2(x) + \cdots + f_i(x)]\,\ell(I) = \sum_{j=1}^{N}\left(\sum_{\sigma_j} f_j(x)\ell(I)\right) \\
&\leq \sum_{j=1}^{N}\left(\overline{\int_a^b} f_j(x)\,dx + \epsilon 2^{-j}\right) \leq \sum_{j=1}^{\infty}\left(\overline{\int_a^b} f_j(x)\,dx\right) + \epsilon.
\end{aligned}$$

This gives an upper bound for all these sums and, since ϵ is arbitrary, shows that

$$\overline{\int_a^b} tf(x)\,dx \leq \sum_{n=1}^{\infty}\left(\overline{\int_a^b} f_n(x)\,dx\right).$$

As this is true whenever $t < 1$, (2) follows. ■

Integrability Criteria. The theoretical development of the integral depends on the following three criteria relating to a function f defined on an interval $[a, b]$:

- **(Cauchy Criterion)** For each $\epsilon > 0$ a Cousin cover β of $[a, b]$ can be found with the property that

$$\left|\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'}[f(z) - f(z')]\ell(I \cap I')\right| < \epsilon$$

  for all partitions π and π' of $[a, b]$ contained in β.
- **(McShane's Criterion)** For each $\epsilon > 0$ a Cousin cover β of $[a, b]$ can be found with the property that

$$\sum_{(I,z)\in\pi}\ \sum_{(I',z')\in\pi'}\left|f(z) - f(z')\right|\ell(I \cap I') < \epsilon$$

  for all partitions π and π' of $[a, b]$ contained in β.

- **(Henstock's Criterion)** There is a function F (called the *indefinite integral*) with the property that for each $\epsilon > 0$ there exists a Cousin cover β of $[a, b]$ with the property that

$$\sum_{(I,x)\in\pi} |\Delta F(I) - f(x)\ell(I)| < \epsilon,$$

for every partition π of $[a, b]$ contained in β.

Cauchy's criterion is just the usual expression of a necessary and sufficient condition in order for a limit to exist. It is easily proved to be equivalent to integrability in a familiar manner.

We have already seen Henstock's criterion in Lemma 3. It becomes, in the present setting, a necessary and sufficient condition for f to be integrable. Simultaneously, it establishes that there must be an indefinite integral F associated with any integrable function that stands in this tight relation to f.

McShane's criterion (from [**13**]) is transparently stronger than the Cauchy criterion. All continuous functions (indeed, all bounded, measurable functions) satisfy this criterion. It is easy to see that if f satisfies this property then so too does $|f|$. The designer of the course should know (but keep that knowledge from the students no doubt) that the following assertions are equivalent:

1. f satisfies McShane's criterion on $[a, b]$;
2. f and $|f|$ are integrable on $[a, b]$;
3. f is Lebesgue integrable on $[a, b]$.

Because of the second of these, it is appropriate to say that a function f is *absolutely integrable* on an interval if it satisfies McShane's criterion there. A natural second definition is that f is *nonabsolutely integrable* if f is integrable but $|f|$ is not.

We know that the lion's share of modern analysis is centered on absolutely integrable functions; relatively little light is cast on nonabsolutely integrable functions. There is a similar situation with convergent series: absolutely convergent series have the most robust properties and the most important applications. Even so, nonabsolutely convergent series and integrals belong naturally to our study and should not be avoided.

An Elephant in the Room? It is the fate of all mathematics that, as it gets more general, it becomes more trivial. The connection between a function and its indefinite integral expressed by Henstock's criterion might appear somewhat profound. The identical connection between a function and its derivative expressed in Lemma 3 was completely trivial. So too is this criterion.

While we have restricted the integral concept to a limit of Cauchy sums of the form

$$\sum_{(I,x)\in\pi} f(x)\ell(I),$$

there is an elephant in the room that we have carefully suppressed. The integral is more natural and compelling a concept if applied to real-valued functions h defined on covering relations. Thus an integral $\int_a^b h$ is defined for such a function h in the same way, namely, as the limit of sums

$$\sum_{(I,x)\in\pi} h(I, x).$$

Then there are three rather simple facts that clarify the Henstock criterion:

(1) $\int_c^d h = 0$ for all subintervals $[c, d]$ of $[a, b]$ if and only if $\int_a^b |h| = 0$.

(2) The statement that F is an indefinite integral of f on $[a, b]$ is merely the observation that

$$\int_c^d (\Delta F - f\ell) = 0$$

for all subintervals $[c, d]$ of $[a, b]$.

(3) The statement that F is an indefinite integral of f on $[a, b]$ is, because of (1) and (2), equivalent to the integration statement that

$$\int_a^b |\Delta F - f\ell| = 0.$$

The designer of the course should be aware that a natural exegesis would ultimately lead to this looser approach to what an integral is but that an elementary course would likely be muddied by this extra generality. Having opened up the integral concept, how could we then resist Stieltjes integrals or more exotic devices?

Perhaps even more seductive is the clarity that such an integral brings to the Jordan decomposition theorem. The expression

$$\overline{\int_a^b} |\Delta f|$$

is precisely the total variation of f on $[a, b]$. If f is continuous and has bounded variation, then two easy identities,

$$\int_a^b \Delta f = \int_a^b [\Delta f]^+ - \int_a^b [\Delta f]^-$$

and

$$\int_a^b |\Delta f| = \int_a^b [\Delta f]^+ + \int_a^b [\Delta f]^-,$$

contain all one needs to know about the Jordan decomposition.

6. MEASURE THEORY. Our goal has not been to suppress the measure theory, but rather to refrain from advancing to techniques that are beyond the scope of the covering lemmas. If we do inject a smattering of measure theory we can introduce ideas that are fundamental to later courses, as well as state and prove significant facts about the derivative and integral, all at this elementary level.

We can begin by allowing a measure theory for open sets and compact sets. This does not exceed our reach since only elementary covering lemmas are needed.

Measure Theory of Open Sets. The *measure* $\mathcal{L}(G)$ of an arbitrary open subset G of $\mathbb{R}$ is defined as follows:

1. $\mathcal{L}(\emptyset) = 0$;
2. $\mathcal{L}((a, b)) = b - a$;
3. $\mathcal{L}(G) = \sum_i (b_i - a_i)$, where (a_i, b_i) are the component intervals of G.

The following lemma asserts the most basic of the measure properties for open sets. While most professional mathematicians can hardly resist an appeal to the Heine-Borel theorem to prove this, we remain with our stated goal of unifying as much as possible by applying covering arguments.

Lemma 5. *If G and G_n $(n = 1, 2, \ldots)$ are open subsets of $\mathbb{R}$ for which $G \subset \bigcup_{n=1}^{\infty} G_n$, then $\mathcal{L}(G) \le \sum_{n=1}^{\infty} \mathcal{L}(G_n)$.*

Proof. Let $\{(a_k, b_k)\}$ be the collection of component intervals of G, fix an integer N no larger than the number of such intervals, and consider the compact set

$$K = \bigcup_{k=1}^{N} [c_k, d_k] \subset \bigcup_{n=1}^{\infty} G_n$$

where c_k and d_k $(k = 1, 2, \ldots, N)$ are arbitrary numbers such that $a_k < c_k < d_k < b_k$. The most natural covering relation in this situation is

$$\beta = \{(I, x) : \ x \in I \text{ and } I \subset G_n \text{ for some } n\}.$$

Checking that β is a quasi-Cousin cover of K, we use Lemma 2 to extract a subpartition π from β for which

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

We collect the intervals I with (I, x) in π that are subsets of a particular set G_n and observe that the total length of these intervals I cannot exceed $\mathcal{L}(G_n)$. There are only a finite number of pairs (I, x) in π to handle, so we see that

$$\sum_{k=1}^{N} (d_k - c_k) \le \sum_{(I,x)\in\pi} \ell(I) \le \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

As this is true for all choices of $[c_k, d_k]$, we conclude that

$$\sum_{k=1}^{N} (b_k - a_k) \le \sum_{n=1}^{\infty} \mathcal{L}(G_n).$$

Since this is true for all relevant N, the inequality of the lemma follows. ■

Measure of Compact Sets. One step further and we can introduce the measure for compact sets: For an arbitrary compact subset K of $\mathbb{R}$ define

$$\mathcal{L}(K) = \inf\{\mathcal{L}(G) : G \text{ open and } G \supset K\}.$$

Measure theory seems to be creeping in to the presentation, but since we have not progressed beyond compact sets, we are no more advanced in our methods than Peano and Jordan were in their late nineteenth century theory of content. We have also avoided their blunder: failing to distinguish between the measure of an open set and the measure of its closure.

While it need play no role in the theory under discussion, it would be interesting for the student to see that the measure $\mathcal{L}$ does have close connections with the integral, even at this early stage.

Lemma 6. *If K is a compact subset of an interval $[a,b]$, then its characteristic function χ_K is integrable on $[a,b]$ and*

$$\mathcal{L}(K) = \int_a^b \chi_K(x)\,dx.$$

Proof. Let G_1 be the open set complementary to K in $\mathbb{R}$ and consider the covering relation

$$\beta_1 = \{(I,x) : x \in K \text{ and } I \subset [a,b] \text{ or } x \notin K \text{ and } I \subset G_1\}.$$

This is a Cousin cover of $[a,b]$. Let π be a partition of $[a,b]$ in β_1. Note that for any (I,x) in π either $x \in K$ and $\chi_K(x) = 1$ or $x \notin K$, $\chi_K(x) = 0$, and $I \cap K = \emptyset$. Set $\pi[K] = \{(I,x) \in \pi : x \in K\}$. Thus

$$\bigcup_{(I,x)\in\pi[K]} I \supset K,$$

whence

$$\sum_{(I,x)\in\pi} \chi_K(x)\ell(I) = \sum_{(I,x)\in\pi[K]} \ell(I) \geq \mathcal{L}(K).$$

This gives a lower bound for the lower integral:

$$\mathcal{L}(K) \leq \underline{\int_a^b} \chi_K(x)\,dx. \tag{4}$$

In the other direction take an arbitrary open set G_2 that contains K and consider the covering relation

$$\beta_2 = \{(I,x) : x \in K \text{ and } I \subset G_2 \text{ or } x \notin K \text{ and } I \subset [a,b]\}.$$

This, too, is a Cousin cover of $[a,b]$. If π is a partition of $[a,b]$ in β_2, observe that for any (I,x) in π either $x \in K$, $\chi_K(x) = 1$, and $I \subset G_2$, or $x \notin K$ and $\chi_K(x) = 0$. Thus

$$\sum_{(I,x)\in\pi} \chi_K(x)\ell(I) = \sum_{(I,x)\in\pi[K]} \ell(I) \leq \mathcal{L}(G_2).$$

This yields an upper bound of $\mathcal{L}(G_2)$ for the upper integral. Since G_2 is an arbitrary open set containing K, we infer that

$$\mathcal{L}(K) \geq \overline{\int_a^b} \chi_K(x)\,dx. \tag{5}$$

Together, (4) and (5) complete the proof. ∎

Null Sets and Null Functions. A subset E of $\mathbb{R}$ is said to be a *null set* if for every $\epsilon > 0$ there is an open set G containing E with $\mathcal{L}(G) < \epsilon$. We need only one simple feature of null sets. The proof is little more than an application of the identity $\sum_{n=1}^{\infty} \epsilon 2^{-n} = \epsilon$ and Lemma 5.

Lemma 7. *If $\{E_n\}$ is a sequence of null sets and $E \subset \bigcup_{n=1}^{\infty} E_n$, then E is also a null set.*

A function $f : \mathbb{R} \to \mathbb{R}$ is a *null function* if $\{x : f(x) \neq 0\}$ is a null set. That null functions are integrable is an essential feature of Lebesgue's theory of integration, easily included in our elementary course.

Theorem 2. *Every null function is integrable on any compact interval, and the value of the integral is zero.*

Proof. The student can be given the case where f is bounded as an exercise along with a hint to mimic the argument in Example 5. The extension to unbounded functions uses a device familiar in measure theory. Fix an interval $[a, b]$. For $m = 1, 2, 3, \ldots$ write

$$E_m = \{x \in [a, b] : m - 1 < |f(x)| \leq m\}.$$

Each set E_m is a null set, and f is bounded on each. An application of Theorem 1 reveals that

$$\overline{\int_a^b} |f(x)|\, dx \leq \sum_{i=1}^{\infty} \overline{\int_a^b} |\chi_{E_i}(x) f(x)|\, dx. \tag{6}$$

As each integrand in the sum on the right is a bounded null function, it follows that each term in the sum is zero. From this we easily conclude that f is integrable on $[a, b]$ with a zero integral. ■

In many presentations of the Henstock-Kurzweil theory some odd pride is taken in the fact that so much can be done (for example this lemma) without introducing the tools of measure theory. In fact, however, this proof does implicitly use some recognizable tools of the measure theory: the splitting up of the set $\{x : f(x) \neq 0\}$ into the sequence $\{E_m\}$, followed by computations similar to outer measure estimates. We are beginning to introduce measure theoretical tools without acknowledging the fact. Inequality (6) should really be expressed as

$$\overline{\int_a^b} |f(x)|\, dx \leq \sum_{i=1}^{\infty} \overline{\int_a^b} |\chi_{E_i}(x) f(x)|\, dx \leq \sum_{i=1}^{\infty} i\mathcal{L}(E_i) = 0, \tag{7}$$

where $\mathcal{L}$ now signifies Lebesgue outer measure. We offer the student no service by burying the concepts deep in the proof and not exposing the underlying measure theoretic facts that actually make the proof more transparent. The moral we should take for our elementary course is to launch the study of measure theory when it is appropriate and otherwise to hold off on obtaining the most general results.

7. THE LEBESGUE DIFFERENTIATION THEOREM.

The Lebesgue differentiation theorem is commonly considered outside the scope of an elementary course, usually waiting for the arrival of the Vitali covering theorem. A proof of this theorem has been given by John Hagood [**3**] that uses the basics of measure theory but replaces an appeal to the Vitali theorem with a simple covering lemma. For convenience we reproduce that here.

Theorem 3. *If $f : \mathbb{R} \to \mathbb{R}$ is a continuous, nondecreasing function, then f has a derivative everywhere outside of a null set.*

Proof. We prove first that the set of points at which f has no right-hand derivative is a null set. We use the notation

$$\overline{D}^{+} F(x) = \lim_{t\to 0+} \sup \left\{ \frac{F(x+s) - F(x)}{s} : 0 < s < t \right\}$$

for the upper right-hand derivative (Dini derivative) of F at x with a similar definition for the lower right-hand derivative $\underline{D}^{+} F(x)$. Recall that F has a right-hand derivative at a point x precisely when these two derivatives agree and are finite.

Thus examination of the set of points at which f has no right-hand derivative leads to an analysis of the set where the upper right Dini derivative exceeds the lower right Dini derivative:

$$E = \{x \in \mathbb{R} : \underline{D}^{+} F(x) < \overline{D}^{+} F(x)\}.$$

This set is, in turn, the countable union of the collection of sets

$$E_{pq} = \{x \in \mathbb{R} : \underline{D}^{+} f(x) < p < q < \overline{D}^{+} F(x)\}$$

taken over all rational numbers p and q with $p < q$. If each of these is a null set then Lemma 7 allows to infer that E is a null set. The only further fact from measure theory we require is the key observation that, in order to verify that E_{pq} is a null set, it is sufficient to show that all its compact subsets are null. (This would require checking for the measurability of Dini derivatives of continuous functions.)

The keys to the proof are the following two growth lemmas, familiar enough, but stated only for compact sets because of our self-imposed limitation of using only simple covering arguments:

Lemma 8. *If $f : \mathbb{R} \to \mathbb{R}$ is a continuous, nondecreasing function and K a compact subset of $\mathbb{R}$ with the property that $\underline{D}^{+} f(x) < p$ for each x in K, then*

$$\mathcal{L}(f(K)) \le p\mathcal{L}(K). \tag{8}$$

Lemma 9. *If $f : \mathbb{R} \to \mathbb{R}$ is a continuous, nondecreasing function and K is a compact subset of $\mathbb{R}$ with the property that $\overline{D}^{+} f(x) > q$ for each x in K, then*

$$\mathcal{L}(f(K)) \ge q\mathcal{L}(K). \tag{9}$$

To prove these lemmas we exploit the geometry of the derivative, expressed in the following two covering relations

$$\beta_p = \left\{ (I, z) : z \in K \cap I,\ \frac{\Delta f(I)}{\ell(I)} < p \right\}$$

and

$$\beta_q = \left\{ (I, z) : z \in K \cap I, \frac{\Delta f(I)}{\ell(I)} > q \right\}.$$

These are not precisely quasi-Cousin covers of K, but they do satisfy the slightly weaker condition expressed in Exercise 2. To verify these local properties, we merely use the fact that $\underline{D}^+ f(x) < p$ (for the first lemma) or $\overline{D}^+ f(x) > q$ (for the second lemma) at each point x of K and make use of the continuity of f.

For Lemma 8, let G be an arbitrary open set containing K. Prune out unnecessary parts of β_p by defining

$$\beta'_p = \{(I, x) \in \beta_p : I \subset G\}.$$

This does not change the covering properties so we can invoke Exercise 2. Thus there is a subpartition π contained in β'_p with the property that

$$K \subset \bigcup_{(I,x)\in\pi} I.$$

Using this, we compute

$$\mathcal{L}(f(K)) \leq \sum_{(I,x)\in\pi} \Delta f(I) \leq \sum_{(I,x)\in\pi} p\ell(I) \leq p\mathcal{L}(G).$$

Since this is true for all such open sets G, we deduce (8).

To establish Lemma 9, consider an arbitrary open set G containing $f(K)$ and remove unnecessary parts of β_q by passing to

$$\beta'_q = \{(I, x) \in \beta_q : I \subset f^{-1}(G)\}.$$

This does not change the covering properties and again Exercise 2 supplies a subpartition π contained in β'_q covering K that leads to the computation

$$q\mathcal{L}(K) \leq \sum_{(I,x)\in\pi} q\ell(I) \leq \sum_{(I,x)\in\pi} \Delta f(I) \leq \mathcal{L}(G).$$

Since this holds for all such open sets G, we infer the correctness of (9).

The two growth lemmas are used to check that each compact subset of either of the following sets is a null set:

$$\{x \in \mathbb{R} : \underline{D}^+ f(x) < p < q < \overline{D}^+ F(x)\} \tag{10}$$

and

$$\{x \in \mathbb{R} : \overline{D}^+ f(x) = \infty\}. \tag{11}$$

Every compact subset K of the first set (10) must satisfy $\mathcal{L}(f(K)) = \mathcal{L}(K) = 0$ because of properties (8) and (9). Also every compact subset K of the second set (11) must satisfy

$$q\mathcal{L}(K) \leq \mathcal{L}(f(K)) \leq f(b) - f(a)$$

for all positive values of q, as a result of property (9). But this can hold only if $\mathcal{L}(K) = 0$.

We are now in a position to complete the proof of Theorem 3. We know that, except for points in a null set (say N_1), the function f has a finite right-hand derivative everywhere in $\mathbb{R}$. Applying that observation to the function $-f(-x)$ shows that, except for points in a null set (say N_2) the function f has a finite left-hand derivative everywhere in $\mathbb{R}$. Thus f has both a left and a right derivative at every point of $\mathbb{R}$ not in the null set $N_1 \cup N_2$. An old and elementary theorem of Dini from the late nineteenth century finishes off the proof: it asserts that left-hand and right-hand derivatives can differ only on a countable set. ■

8. DIFFERENTIATION OF THE INTEGRAL. If f is integrable on an interval $[a, b]$ and F is an indefinite integral of f there (in the sense of the Henstock criterion) then

$$F'(x) = f(x)$$

for every x in $[a, b]$ excepting a null set. For the Riemann integral it is common to prove this at points of continuity, since except for a null set every point for a Riemann integrable function is a point of continuity. For the Lebesgue integral this requires some nontrivial measure theory.

A nearly elementary proof, however, is possible provided that we are willing to add the hypothesis that the function f is measurable and to invoke the same measure theory basics as just used in Section 7. We might be tempted to advertise this as a proof that makes no use of the Vitali covering theorem, but there are aspects of Vitali covers in the details. What we have done is make the quasi-Cousin covering lemma handle the job that, in a more general presentation, the Vitali covering theorem would tackle.

Theorem 4. *If $f : [a, b] \to \mathbb{R}$ is a measurable function that is integrable on an interval $[a, b]$ and F is an indefinite integral on $[a, b]$, then $F'(x) = f(x)$ for every x in $[a, b]$ outside of a null set.*

Proof. We show that $\overline{D}^{+}F(x) = \underline{D}^{+}F(x) = f(x)$ for each x in $[a, b]$ outside of a null set. The set of points where $\underline{D}^{+}F(x) < f(x)$ can be analysed (as before) by considering the sets

$$E_{pq} = \{x \in [a, b] : \underline{D}^{+}F(x) < p < q < f(x)\}$$

for rational numbers p and q with $p < q$. If each of these sets is a null set, then

$$E = \{x \in [a, b] : \underline{D}^{+}F(x) < f(x)\} \tag{12}$$

is also a null set. As before, we use the principle that if every compact subset of E_{pq} is a null set then E_{pq} is as well. (Usual measure theoretic arguments allow this once it is noted that both f and $\underline{D}^{+}F$ are measurable functions.)

To this end, let $\epsilon > 0$ and select a Cousin cover β_0 of $[a, b]$ with the property that the Henstock criterion in Section 5 is satisfied. Let K be any compact subset of E_{pq}, and let

$$\beta_1 = \{(I, x) : x \in K,\ \Delta F(I) < p\ell(I) < q\ell(I) < f(x)\ell(I)\}.$$

While β_1 falls a little short of being a quasi-Cousin cover of K, we can check that it satisfies the hypotheses of Exercise 2 for the set K by using the inequalities

$$\underline{D}^{+}F(x) < p < q < f(x)$$

and the continuity of F at each point.

The intersection $\beta_0 \cap \beta_1$ also satisfies these same hypotheses. We invoke Exercise 2 to choose a subpartition π from $\beta_0 \cap \beta_1$ covering K that allows the following computation:

$$(q-p)\mathcal{L}(K) \le (q-p) \sum_{(I,x)\in\pi} \ell(I) \le \sum_{(I,x)\in\pi} [q\ell(I) - p\ell(I)]$$
$$\le \sum_{(I,x)\in\pi} |f(x)\ell(I) - \Delta F(I)| < \epsilon.$$

As $\epsilon > 0$ is arbitrary and $q - p > 0$ it follows that $\mathcal{L}(K) = 0$. Since this is true for any compact subset K of E_{pq}, we infer that E_{pq} is a null set. As announced earlier, this shows that the set E in (12) is a null set.

Exactly the same argument, with suitable modifications, shows that the set of points where $\overline{D}^{+} F(x) > f(x)$ is a null set. Consequently

$$\underline{D}^{+} F(x) = \overline{D}^{+} F(x) = f(x)$$

at every point x of $[a, b]$ outside of a null set.

This result can be applied to the functions $f(-x)$ and $F(-x)$ to show that, again outside of a null set,

$$\underline{D}^{-} F(x) = \overline{D}^{-} F(x) = f(x).$$

The theorem follows. ■

9. TAGGED, GAUGED, GAGGED, AND MUGGED. Presentations of the Henstock-Kurzweil integral seem to have settled into a routine language over the years, a language that has done little to promote its use. Originally there was an enthusiastic portrayal of the integral as an obvious generalization of the Riemann integral: just change the δ in the freshman calculus course to a δ that changes from point-to-point. Call δ a "gauge." Call it the "gauge integral." Partitions used to be collections $\{I_i\}$ of subintervals; since they must now assume the form $\{(I_i, x_i)\}$, call them "tagged partitions." The x_i are the "tags." Tagged partitions are "δ-fine."

Every proof has a gauge or several gauges, many tags, and nearly everything is δ-fine. Detailed manipulations of the gauges and tags consume the attention of the reader. Since the gauges are in the driver's seat, most concepts arise from them. Convergence criteria, variational notions, absolute continuity, all of the most vital concepts are gauged and tagged to death. For the reader unwilling to enter into this world, there are few rewards and little encouragement. On how many of my students would I be willing to impose this tagged and gauged language knowing that, when they escape into graduate analysis, they will never need it again?

In its place we have substituted the language of covering relations. Covering relations have many merits as a fundamental concept. Proofs, freed from a need to relate everything to a gauge, are much more transparent. The Cousin covering lemma is a reasonable centerpiece for a first analysis course. It has the special merit that it can be used for all of the compactness arguments in that course and serve, at the same time, as the central concept used to define the integral. Will the professional reader be any more pleased now or will he feel equally assaulted and mugged as by the tagged and gauged crowd?

Several generations have now shown little interest in moving these ideas into the undergraduate curriculum. Is the Riemann integral here to stay? Will any better notation or presentation change things?

10. FINAL REMARKS. Let me close with a quotation from a 1981 *Mathematical Reviews* report on an earlier work [**11**] that attempted an analysis text along the lines suggested here:

> Proofs and methods of notation are very detailed, occasionally almost pedantic and rather lengthy. The circumvention of the compactness arguments via "Cousin's theorem" seems artificial to the reviewer; many of the familiar inferences thus appear rather awkward here. It is probably—and with good reason, in the reviewer's opinion—rather unusual to introduce the Perron integral instead of the Lebesgue integral into class lectures; and for the students who were given a concise approach to the Lebesgue integral this book would naturally not be very suitable for supplementary reading.

It is easy to imagine that a similar book published today would receive an identically dismissive treatment. (A later version of this same author's analysis text was given an enthusiastic write-up in *Mathematical Reviews* by Bob Bartle, but we already know that Bartle was a sympathetic audience.)

These ideas deserve to be aired and considered by future writers of elementary real analysis texts. There are many viewpoints possible, and many, very different, treatments are possible. A recent posting by Smithee [**14**] offers one possibility. An alternative solution has been available in Zakon's on-line text [**16**] where all constructive integration techniques have been dropped in favor of a Newton-type integral which includes most applications that one would reasonably require at this level. (The key lemma that would justify this integral is a simple application of the covering arguments of Section 3.)

In spite of a detailed knowledge of the Henstock-Kurzweil integral for many years and a long friendship with Ralph Henstock, I have always been reluctant to promote the integral for the undergraduate curriculum. For one thing it was clear from the outset that there would be resistance, and I had no crusading inclinations. More importantly, I have never agreed that this integral should be an excuse to avoid measure theory. There is much to say about it, but one obvious point should be made. The integral is purely formal, not constructive. While it includes Lebesgue's integral, no real understanding of the integral and technical facility in using it are possible without incorporating all of Lebesgue's methods. The true advantage possessed by this newer theory is that it allows for a better and clearer presentation of the standard material, not a dilution.

The same goes for the nonabsolutely integrable functions. Quite transparently the integral is *formally* simpler than Denjoy's totalization process, but were Denjoy here to defend himself (and he would!), he would insist that to understand this integral requires an understanding of the extension processes he used and an account of the nature of the integral along constructive lines using transfinite ordinals.

DEDICATION.

Dedicated to Robert G. Bartle (1927–2003).

REFERENCES

1. R. G. Bartle, *A Modern Theory of Integration*, Graduate Studies in Math., no. 32, American Mathematical Society, Providence, 2001.
2. ———, Return to the Riemann integral, this MONTHLY **103** (1996) 625–632.
3. J. Hagood, The Lebesgue differentiation theorem via nonoverlapping interval covers, *Real Anal. Exch.* **29** (2003–04) 953–956.
4. J. Hagood and B. S. Thomson, Recovering a function from a Dini derivative, this MONTHLY [to appear].
5. R. Henstock, The efficiency of convergence factors for functions of a continuous real variable, *J. London Math. Soc.* **30** (1955) 273–286.
6. ———, A Riemann-type integral of Lebesgue power, *Canad. J. Math.* **20** (1968) 79–87.
7. J. Kurzweil, Generalized ordinary differential equations and continuous dependence on a parameter, *Czechoslovak Math. J.* **7** (1957) 418–449.
8. S. Leader, What is a differential? A new answer from the generalized Riemann integral, this MONTHLY **93** (1986) 348–356.

9. ———, *The Kurzweil-Henstock Integral and Its Differentials. A Unified Theory of Integration on* $\mathbb{R}$ *and* $\mathbb{R}^n$, Marcel Dekker, New York, 2001.
10. H. Lebesgue, *Leçons sur L'Intégration*, Chelsea, New York, 1973; reprint of the 1928 Paris 2nd ed.
11. J. Mawhin, *Introduction á l'analyse,* 2nd. Cabay Libraire-Éditeur S.A., Louvain-la-Neuve, France, 1981.
12. E. J. McShane, A unified theory of integration, this MONTHLY **80** (1973) 349–359.
13. ———, *Unified Integration*, Academic, New York, 1983.
14. A. Smithee, *The Integral Calculus*, available at `http://www.classicalrealanalysis.com`.
15. B. S. Thomson, On full covering properties, *Real Anal. Exchange* **6** (1980/81) 77–93.
16. E. Zakon, *Mathematical Analysis*, available at `http://www.trillia.com`.

BRIAN S. THOMSON received his undergraduate education at the University of Toronto and his graduate degrees at the University of Waterloo. His first academic position was in Waterloo, following which he moved in 1968 to Simon Fraser University where he remains, now as Professor Emeritus. His research interest is in classical real analysis and he is a co-author of two real analysis textbooks. Currently he serves on the editorial boards of the Real Analysis Exchange and the Journal of Mathematical Analysis and Applications.
Mathematics Department, Simon Fraser University, B.C., Canada
thomson@cs.sfu.ca

Young Drumgoole, or Drum as they called him, was brought up with the same narrowness of view which his mother had so painstakingly implanted in his sister. From the beginning his father had fondly hoped that the youth would carry on the war-like strain and had envisioned the proud day when he should be a candidate for West Point. And eventually the proud day came, and Drum was sent away into the very heart of the enemy's country with many fond flourishes of fatherly advice and admonition. But it was all to no purpose. He never saw the conclusion of his first term at the great academy on the Hudson. He was a casualty to the rapid fire of trigonometry. He was cut down the first charge. He never even heard the booming of the heavy guns upon the distant front of calculus.

After this, of course, there was nothing left for him to do except to go to Charlottesville and enroll himself among the princelings of the blood at the University of Virginia. Any other alternative was clearly impossible. A gentleman could still attend the United States Military Academy without dishonor, for Lee himself had been a West Point man, but to submit his person to the Yankee degradations of Harvard, Yale, or Princeton was, in the eyes of the Colonel and his wife, unthinkable.

So Drum was sent to Charlottesville as the next-best thing to West Point. Of his life there, there is little to record save that he finally scraped through and learned to "hold his liquor like a gentleman"—which apparently has always been one of the stiffest requirements of the curriculum at that famous university. At length Drum came home again wearing a small blonde mustache, and was instantly appointed a Major in his father's celebrated corps, and second in command, the appointment also carrying with it an instructorship in mathematics, trigonometry and calculus included.

Thomas Wolfe, *The Hills Beyond*
New American Library, New York & Toronto, 1941, pp. 216–217
—Submitted by Robert Haas, Cleveland Heights, OH