THE EVOLUTION OF . . .

Edited by **Abe Shenitzer**

Mathematics, York University, North York, Ontario M3J 1P3, Canada

An English major may or may not be a novelist or a poet, but would undoubtedly be expected to be able to evaluate a novel or a poem. The term "English major" implies some historical, philosophical, and evaluative training and competence. It is sad but true that the term "mathematician" does not imply corresponding training and competence.

Integration of the narrowly mathematical and historical, philosophical and critical aspects of our discipline is bound to make it more meaningful not only to those who identify themselves as mathematicians but also to those who have no more than a tangential interest in the subject.

To promote such integration, and thus encourage an approach to mathematics that emphasizes its meaning and significance, the Monthly will publish every two months an article of 2–5 pages under the generic title "The evolution of..." The core of such an article will be an account of important mainstream mathematics. The essay that follows exemplifies the kind of material, and the approach, we have in mind.

While we prefer original articles, we will also publish translations or adaptations of appropriate articles in the public domain.

Abe Shenitzer

The Evolution of Integration

A. Shenitzer and J. Steprāns

THE GREEK PERIOD. The Greek problem underlying integration is the *quadrature problem*: Given a plane figure, construct a square of equal area.

It is easy to solve the quadrature problem for a polygon, a figure with rectilinear boundary. The first quadrature of a figure with curvilinear boundary was achieved by Hippocrates in the fifth century B.C. Hippocrates showed that the area of the lunule in FIGURE 1 (that is, the figure bounded by one-half of a circle of radius 1 and one-quarter of a circle of radius $\sqrt{2}$) is equal to the area of the unit square B.

Hippocrates managed to square two other lunules.*

In the third century B.C. Archimedes effected the quadrature of a parabolic segment. He showed that its area is $\frac{4}{3}\Delta$, where Δ is the triangle of maximal area inscribed in the parabolic segment.

Archimedes effected a number of other quadratures (and cubatures). Some of his quadratures involved inventive constructions but most relied on the technique of wedging an area between ever closer upper and lower approximating sums.

*Two more quadrable lunules were found by T. Clausen in the 19th century. In the 20th century, two Russian algebraists proved (independently) that these five lunules are the only quadrable ones.

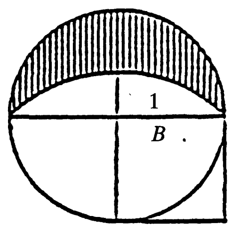


Figure 1

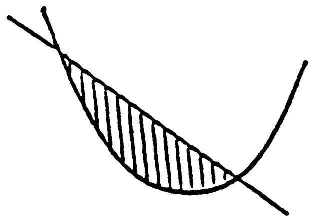


Figure 2

Analogs of such sums are a key element of the definition of the Darboux integral (a variant of the Riemann integral introduced by Darboux in the 19*th century) as well as of quadrature programs for computers.* We illustrate both of Archimedes' approaches next.

Consider FIGURE 3. Here the hypotenuse AB of the right triangle OAB is tangent to the spiral at A. It then turns out that the side AB is equal to the

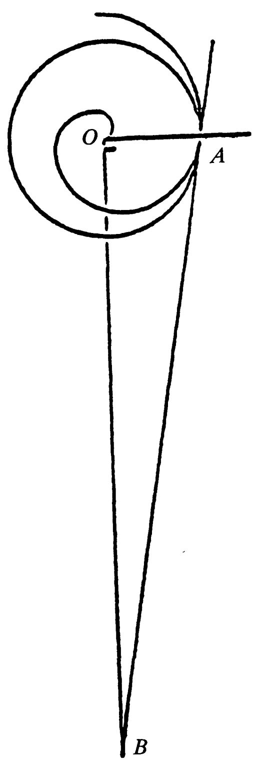
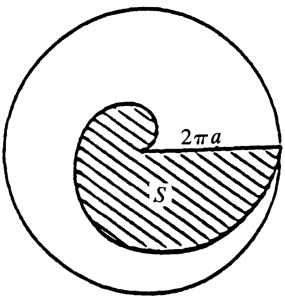


Figure 3

circumference of the circle with radius OA. (This is a special case of Archimedes' rectification of circular arcs by using tangents to spirals.) Since he knew that the area of a circle is half the product of its circumference by its radius, we can say that *Archimedes used (a tangent to) a spiral to rectify a circle and square its area.* Their brilliance notwithstanding, such constructions have been reduced to historical footnotes because they failed to yield general methods.

FIGURE 4 shows a turn of Archimedes' spiral $r = a\theta$ and the associated circle of radius $2\pi a$, and thus of area $K = 4\pi^3 a^2$. To compute the area S of the turn of the spiral in FIGURE 4 Archimedes approximates it from below and above by unions of circular sectors indicated in FIGURE 5.



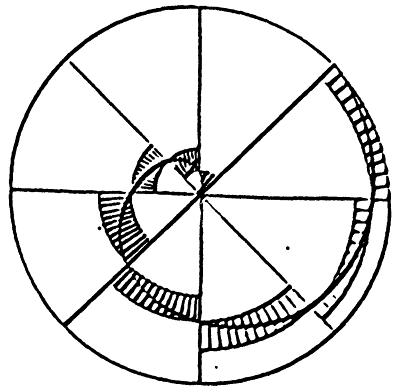


Figure 4 Figure 5

The areas of these approximating figures are, respectively,

$$S'_n = \frac{4\pi^3}{n^3}\left[1^2 + 2^2 + \cdots + (n-1)^2\right] = \frac{2\pi^3 a^2 (n-1)(2n-1)}{3n^2}$$

and

$$S''_n = \frac{4\pi^3}{n^3}\left[1^2 + 2^2 + \cdots + n^2\right] = \frac{2\pi^3 a^2 (n+1)(2n+1)}{3n^2}.$$

It is not difficult to see that

$$S'_n < \frac{4}{3}\pi^3 a^2 < S''_n$$

for all n. This double inequality can be rewritten as

$$S'_n < \frac{K}{3} < S''_n$$

for all n. Obviously,

$$S'_n < S < S''_n$$

for all n. To prove that $S = (K/3)$ Archimedes shows that $S''_n - S'_n = (4\pi^3 a^2/3n^2)$ and is thus small for large n. He can now show that the assumption $S \neq K/3$ leads to a contradiction and can conclude that $S = (K/3)$.

While Archimedes makes no explicit use of limits, he relies on the "method of exhaustion," and, in modern terms, the final part of the argument in a proof involving the method of exhaustion (in the above example it is disproving $S \neq K/3$) amounts to proving the uniqueness of the limit of a Cauchy sequence.

CONTINUATION IN THE 17TH CENTURY. Using nonrigorous infinitesimal techniques (rather than rigorous algebraic methods of the kind used by Archimedes) Cavalieri (1598–1647) managed to compute (what we now write as) $\int_0^1 x^k\,dx$ for $k = 1, 2, \dots, 9$. His chief difficulty was the evaluation of $1^k + \cdots + n^k$. In about 1650 Fermat evaluated $\int_0^a x^{p/q}\,dx$ by means of a brilliant yet simple computation. Further progress was due to Torricelli, Wallace, and Pascal. In particular, Pascal interpreted Cavalieri's "sum of lines" (the equivalent of area) as a sum of infinitesimal rectangles.

If we combine Fermat's result with Cavalieri's understanding of the linearity of the definite integral (*our* terminology!) then we see that by the middle of the 17th century one could evaluate $\int_a^b P(x)\,dx$, $P(x)$ a "polynomial" with rational exponents.

In 1647 Gregory St. Vincent made a discovery that linked Napier's logarithm function and the area under the hyperbola $xy = 1$. This connection is now expressed as $\log_e(x) = \int_1^x (dt/t)$.

Newton and Leibniz invented the calculus and made it into a tool with countless applications but neither gave what we would call a rigorous definition of a definite integral (or saw the need for such a definition). Such concerns became dominant in the 19th century.

FROM CAUCHY TO LEBESGUE. The first rigorous definition of a definite integral was given by Cauchy in the 1820s. Cauchy dealt with continuous functions. In view of the importance of Fourier series whose coefficients are given by integrals it was necessary to define the integral for more general functions. This was first done by Riemann. The limitations of the Riemann integral were remedied at the beginning of the 20th century by Lebesgue. An explanation follows.

With each theory of integration there is associated a theory of measure. Specifically, if f is a function on a set E and $f = f^+ - f^-$ (recall that $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$) then $\int_E f$ is defined as the difference $\int_E f^+ - \int_E f^-$ of the measures $\int_E f^+$ and $\int_E f^-$ of the ordinate sets of the nonnegative functions f^+ and f^- respectively.

The measure underlying the Riemann integral is Jordan measure and the measure underlying the Lebesgue integral is Lebesgue measure. How do they differ? In what way is one "better" than the other?

Consider the simple case of the ordinate set M of a bounded, nonnegative function f on an interval, $0 \le f(x) \le c$ for x in $[a, b]$. The Jordan measure of M is the common value, if any, of the outer and inner Jordan measures of M. The outer Jordan measure of M is the glb of the areas of the coverings of M consisting of *finite* unions of rectangles. The inner measure of M is the difference between the area $C(b - a)$ of the rectangle S with base $[a, b]$ and height C and the outer measure of the complement of M in S. Lebesgue replaced the word "finite" in the Jordan definition of the measure of a subset of S by "countable." This increased greatly the number of measurable subsets of S and led to a theory of integration far more comprehensive and mathematically flexible than Riemann's.

THE HK-INTEGRAL. Surprisingly, Henstock (in 1955) and Kurzweil (in 1957) came up with a new version of the Riemann integral—call it the HK-integral (see [7])—that is "as good as" the Lebesgue integral! Its definition and main characteristics follow (see [7]):

Definition: A *tagged division* of $[a, b]$ given by a finite ordered set $a = x_0 < x_1 < \cdots < x_n = b$ of points, together with a collection of *tags* z_i such that $x_{i-1} \le z_i \le$

x_i for $i = 1, \ldots, n$. We denote a tagged division by $D(x_i, z_i)$ and the corresponding Riemann sum by

$$S(D(x_i, z_i)) := \sum_{i=1}^{n} f(z_i)(x_i - x_{i-1}).$$

A *gauge* on $[a, b]$ is a function δ defined on $[a, b]$ such that $\delta(x) > 0$ for all $x \in [a, b]$. An important example of a gauge is a constant function. If δ is any gauge on $[a, b]$, we say that a tagged division $D(x_i, z_i)$ is *δ-fine* in case that $[x_{i-1}, x_i] \subseteq [z_i - \delta(z_i), z_i + \delta(z_i)]$; that is, in case $z_i - \delta(z_i) \leq x_{i-1} \leq z_i \leq x_i \leq z_i + \delta(z_i)$ for all $i = 1, 2, \ldots, n$. Finally, we say that the number A is an *HK-integral* of f if, for every $\varepsilon > 0$, there exists a gauge δ_ε such that if $D(x_i, z_i)$ is any tagged division of $[a, b]$ that is δ_ε-fine, then we have

$$|S(D(x_i, z_i)) - A| < \varepsilon.$$

It turns out that "the *HK*-integral of a function is uniquely defined when it exists and that a function is Riemann integrable if and only if the gauge δ_ε can be chosen to be constant." More importantly, "every Lebesgue integrable function is *HK*-integrable with the same value."

THERE IS NO PERFECT INTEGRAL. While in the eyes of some mathematicians the Lebesgue integral was the final answer to the difficulties associated with integration, there were others who were not willing to give up the search for the *perfect* integral, one which would make all functions integrable. Because Lebesgue's construction had shown that the key to a comprehensive theory of integration was the construction of an appropriate measure, the search now focussed on finding a total measure on $\mathbb{R}$, that is, one which assigns a measure to each subset of the real numbers.

Vitali [**6**] showed that a total measure on the reals cannot be countably additive *and* translation invariant. This being so, it is natural to ask which of these properties should be retained. This decision is, of course, somewhat arbitrary. While retaining translation invariance leads to some fascinating group theory and the Banach-Hausdorff-Tarski Paradox, we will consider what happens if countable additivity is retained instead.

In 1930 S. Ulam [**1**] showed that there is no such measure on ω_1, ω_2 or on any cardinal[1] which is the successor of some other cardinal. Ulam's proof was a spectacular advance in that it did not rely on any of the geometric assumptions, such as translation invariance, on which earlier proofs of the existence of non-measurable sets had relied.

By Ulam's theorem, the existence of a countably additive measure on $\mathbb{R}$ that measures all of its subsets implies that $2^{\aleph_0}$ is not the successor of any other cardinal, that is, it is a limit cardinal. By arguing a bit more carefully one can show that there must exist some limit cardinal $\lambda \leq 2^{\aleph_0}$ which is not the union of fewer than λ sets of size less than λ. The existence of such a cardinal has a profound influence on set theory.

In order to understand this influence, it is necessary to recall (a consequence of) Gödel's second incompleteness theorem which says that set theory can not prove its own consistency. One way to prove the consistency of a theory is to find a model

[1]Cantor introduced the notation ω to represent the next ordinal after the integers and it is still favored by set theorists today. The next cardinal after ω is denoted ω_1 and so on.

of that theory, that is, a mathematical structure satisfying all of the axioms of that theory. We ask: What are the implications for set theory of the existence of a model of set theory? Recall the procedure for the construction of the hierarchy of sets. One begins with the empty set—call it V_0—and then defines V_k to be the power set of V_{k-1} for each integer $k \geq 1$. This is not the end, though, because one can then define V_ω to be the union of the sets V_k and then define $V_{\omega+1}$ to be the power set of V_ω. If one continues this as far as possible and takes the union one gets a model of set theory—or, at least, what would be a model of set theory if it were a set and not a proper class.

How soon, if ever, does this construction process lead to a model of set theory? It turns out that many of the axioms of set theory are satisfied at early stages of the construction. For example the axiom of infinity is satisfied as soon as a single infinite set is included and this is already true of $V_{\omega+1}$. The power set axiom is satisfied at any limit stage because any set which occurs, occurs at a stage before the limit and so all of its subsets are added at the very next stage. The power set itself is therefore added in no more than two stages and, in any case, before the limit. For similar reasons, the pairing axiom is also satisfied at all limit stages. Well-foundedness and comprehension are also easy to deal with.

The problematic axiom is the axiom of replacement, which says that the range of any function defined by a formula is a set. It has already been mentioned that $V_{\omega+\omega}$ will satisfy all of the axioms of set theory except for replacement. Replacement fails because the mapping which takes $2n$ to $\omega + n$ and $2n + 1$ to n is definable by a formula and its domain is ω which belongs to $V_{\omega+1} \subseteq V_{\omega+\omega}$. However, the range of this function is $\omega + \omega$ which does not belong to $V_{\omega+\omega}$. The same argument can be used to show that V_α is a model of set theory if and only if the following holds:

- if $\lambda < \alpha$ then $2^\lambda < \alpha$
- if $\lambda < \alpha$ then any function $F: \lambda \to \alpha$ (defined using only parameters from V_α) has range bounded in α.

Any cardinal satisfying these requirements is known as a *large* or *inaccessible* cardinal. Since the existence of a large cardinal implies that a model of set theory exists, it follows from Gödel's Theorem that it is impossible to prove the existence of inaccessible cardinals.

Ulam's argument shows that if there is a countably additive measure which measures every set of reals then there is a cardinal α which satisfies the second requirement of being an inaccessible cardinal. Such cardinals are known as weakly inaccessible. Another of Gödel's major contributions is the notion of the Constructible Universe, one of whose consequences is that any model of set theory contains a submodel which satisfies the generalized continuum hypothesis. This allows us to conclude that if there is a weakly inaccessible cardinal then, in the Constructible Universe, the weakly inaccessible cardinal is in fact an inaccessible cardinal; this is so because the cardinal arithmetic of this smaller model of set theory easily implies the first requirement for being a large cardinal.

In other words, if there is a countably additive measure which measures every set of reals than set theory is consistent. This and Gödel's theorem show that the existence of a *perfect* integral is not provable. On the other hand, it is conceivable that some day there may be a proof that it is *not* possible to have a perfect integral. The impact of this on set theory would be devastating. It would follow that many of the large cardinals which experts now consider quite innocuous, and which have played an important role in many important independence results, do not exist. While this would not show that set theory itself is inconsistent it would severely shake our faith in the assumption that it is.

* * *

We've told our story but would nevertheless like to tack on the following relevant "postscript":

In what sense does the integral solve the Greek quadrature problem and what is its conceptual significance? A telegraphic answer to these two questions follows.

The integral provides a direct "analytic" solution of the Greek quadrature problem for regions of the form

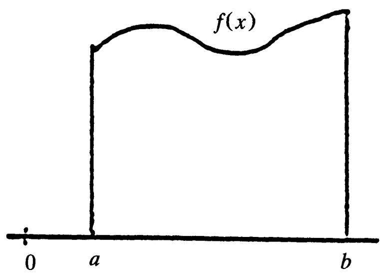


Figure 6

Indeed, the area of the region in the figure is

$$A = \int_a^b f(x)\,dx.$$

If we rewrite this as

$$A = \int_a^b f(x)\,dx = (b-a)\left(\frac{1}{b-a}\int_a^b f(x)\,dx\right),$$

then it is clear that our "integral region" has been replaced by a rectangle of equal area with base $b - a$ and height $(1/(b-a))\int_a^b f(x)\,dx$. The quantity $(1/(b-a))\int_a^b f(x)\,dx$ is the average of the functional values of f on $[a, b]$. *This averaging ability of the integral is the key to its importance in countless applications.*

REFERENCES

1. S. Ulam, Zur Masstheorie in der allgemeinen Mengenleh, *Fund. Math.* 16 (1930), 140–150.
2. C. H. Edwards, Jr. *The Historical Development of the Calculus*, Springer-Verlag, 1979.
3. O. Toeplitz, *The calculus—A Genetic Approach*, University of Chicago Press, 1963.
4. A. Aaboe, *Episodes from the Early History of Mathematics*, the MAA, NML 13.
5. T. Jech, *Set Theory*, Academic Press, New York, 1978.
6. G. Vitali, *Sul problema della mesura dei gruppi di punti di una retta*, Bologna, 1905.
7. R. G. Bartle, review of R. Henstock's *The General Theory of Integration*, BAMS, v. 29, #1, July 1993, pp. 136–139.

Department of Mathematics and Statistics
York University
North York, Ontario
Canada M3J 1P3.