A Historical Gem from Vito Volterra

WILLIAM DUNHAM
Hanover College and
The Ohio State University
Columbus, OH 43210

Sometime during the junior or senior year, most undergraduate mathematics students first examine the theoretical foundations of the calculus. Such an experience—whether it's called "analysis," "advanced calculus," or whatever—introduces the precise definitions of continuity, differentiability, and integrability and establishes the logical relationships among these ideas.

A major goal of this first analysis course is surely to consider some standard examples—perhaps of a "pathological" nature—that reveal the superiority of careful analytic reasoning over mere intuition. One such example is the function defined on $(0,1]$ by

$$g(x)=\begin{cases}1/q & \text{if } x=p/q \text{ in lowest terms}\\ 0 & \text{if } x \text{ is irrational}\end{cases}$$

This function, of course, is continuous at each irrational point in the unit interval and discontinuous at each rational point (see, for instance [**2**, p. 76]). It thus qualifies as extremely pathological, at least to a novice in higher mathematics. By extending this function periodically to the entire real line, we get a function—call it "G"—continuous at each irrational number in $\mathbb{R}$ and discontinuous at each rational in $\mathbb{R}$.

The perceptive student, upon seeing this example, will ask for a function continuous at each *rational* point and discontinuous at each *irrational* one, and the instructor will have to respond that such a function cannot exist.

"Why not?" asks our perceptive, and rather dubious, student.

In reply, the highly trained mathematician may recall his or her graduate work and begin a digression to the Baire Category Theorem, with attendant discussions of nowhere dense sets, first and second category, F_σ's and G_δ's, before finally demonstrating the non-existence of such a function (for example, see [**6**, p. 141]). Of course, it takes quite a while to set up all this sophisticated mathematical machinery, during which time the student probably will have lost interest or graduated.

It may come as a surprise, then, that this question was answered in a short and simple 1881 proof by the brilliant Vito Volterra (1860–1940). Discovered by Volterra when he was still a student at Pisa's Scuola Normale Superiore, his result predates René Baire's groundbreaking category theorem [**1**] by almost two decades yet uses only the relatively unsophisticated notions of "continuous function" and "dense set." As such, it provides a fine example of the advantages of studying the history of mathematics; for not only does it give a glimpse into the past but simultaneously satisfies the classroom needs of the present.

The argument appeared early in Volterra's paper "Alcune osservasioni sulle funzioni punteggiate discontinue" [**8**, pp. 7–8]. Central to this work was the concept of "pointwise discontinuous" functions, i.e., functions whose points of continuity form a dense set. The function $G(x)$ above is pointwise discontinuous, and other such functions had appeared on the scene by the mid-nineteenth century. Bernhard Riemann [**5**, p. 242], for instance, startled his 1854 audience with an example of an integrable function having discontinuities precisely at rationals of the form $m/2n$

where m and $2n$ are relatively prime. To some, this suggested an intimate link between pointwise discontinuity and the highly complicated matter of integrability.

One such mathematician was Hermann Hankel, who in 1870 made a detailed examination of the "punktirt unstetigen," i.e., "pointwise discontinuous," functions. (This rather odd-sounding term was introduced as a contrast to the "total unstetige," i.e., "totally discontinuous," functions, whose points of continuity were not dense.) Hankel gave a proof [3, pp. 89–90] purporting to show that a function is (Riemann) integrable if and only if it is pointwise discontinuous. This conclusion certainly thrust pointwise discontinuous functions into the limelight, and, as Thomas Hawkins observes in his excellent book *Lebesgue's Theory of Integration: Its Origins and Development* [4, p. 30], seemed to identify them as precisely "the functions amenable to mathematical analysis."

However, Hankel's reasoning was flawed, and the error was exposed in 1875 by Oxford professor H. J. S. Smith [7, p. 150]. While Smith agreed that any integrable function must be pointwise discontinuous, his explicit example of a pointwise discontinuous but non-integrable function destroyed the "if and only if" nature of Hankel's proof. Smith had established, of course, that the integrable functions form a *proper* subset of the pointwise discontinuous ones, but his paper was not widely read and its impact was minimal. Consequently, pointwise discontinuous functions occupied an important, although not necessarily well-understood, position in the research of the day. (Hawkins gives a detailed account of the situation in [4, Ch. 2].)

It was in this context that the young Volterra composed his 1881 paper. There, he stated the following key theorem:

THEOREM. *There do not exist pointwise discontinuous functions defined on an interval (a, b) for which the continuity points of one are the discontinuity points of the other, and vice versa.*

Beginning a proof by contradiction, Volterra assumed the existence of two such functions, f and ϕ. For notational ease, we shall let

$$C_f = \{x \in (a,b) | f \text{ is continuous at } x\}.$$

Thus, Volterra's assumption was that the dense sets C_f and C_ϕ partition (a, b) into disjoint subsets.

Let x_0 be any point in C_f and take $\alpha = 1$. Continuity guarantees the existence of a $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset (a, b)$ and $|f(x) - f(x_0)| < 1/2$ for all $x \in (x_0 - \delta, x_0 + \delta)$. We then can choose $a_1 < b_1$ so that $[a_1, b_1]$ is a *closed* subinterval of $(x_0 - \delta, x_0 + \delta)$ and consequently, for any x and y in $[a_1, b_1]$,

$$\begin{aligned}|f(x) - f(y)| &\leqslant |f(x) - f(x_0)| + |f(x_0) - f(y)| \\ &< 1/2 + 1/2 = 1 = \alpha.\end{aligned}$$

Pointwise discontinuity now yields a continuity point of ϕ in the open interval (a_1, b_1), and by the preceding argument there exist $a_1' < b_1'$ with $[a_1', b_1'] \subset (a_1, b_1)$ and with

$$|\phi(x) - \phi(y)| < 1 \text{ for all } x \text{ and } y \text{ in } [a_1', b_1'].$$

To summarize, then, for all x and y in $[a_1', b_1'] \subset (a, b)$,

$$|f(x) - f(y)| < 1 \text{ and } |\phi(x) - \phi(y)| < 1$$

as well.

Volterra next repeated this argument, starting with the open interval (a_1', b_1') and the margin $\alpha = 1/2$, then $\alpha = 1/4$, and generally $\alpha = 1/2^n$. This generates a strictly descending sequence of closed intervals

$$(a, b) \supset [a_1', b_1'] \supset \cdots \supset [a_n', b_n'] \supset \cdots$$

such that, for all x and y in $[a_n', b_n']$, we have both

$$|f(x) - f(y)| < 1/2^n \text{ and } |\phi(x) - \phi(y)| < 1/2^n.$$

(Note how the pointwise discontinuity plays a central role at each step, providing at least one continuity point for f or ϕ in any open subinterval.)

But by the Nested Interval Theorem there exists at least one point A contained in *all* of the closed subintervals above, and thus both f and ϕ are continuous at A. In short, $C_f \cap C_\phi \neq \varnothing$, a contradiction. From this, Volterra concluded that no such functions f and ϕ exist. Q.E.D.

We should observe that Volterra was a bit vague about the intervals $[a_n', b_n']$ being CLOSED, as indeed they must be to guarantee a point in their intersection. Volterra was not alone among nineteenth century mathematicians in this vagueness. Hankel, in the paper cited above, failed to stress this pivotal detail, as did Baire in his proof of the "Category Theorem" that carries his name [**1**, p. 65]. Fortunately, Volterra's argument is easily repaired, as has been done above.

From this simple proof, Volterra then drew two interesting conclusions. The first, answering the question of our perceptive student, was that no function can have the set of rationals as its only points of continuity, for such a function would be pointwise discontinuous with continuity points corresponding to the discontinuity points of the "extended" pathological function G above, a situation whose impossibility he had just demonstrated.

Second, Volterra reasoned that there can be no continuous function ϕ mapping rationals to irrationals and vice versa. For, if such a ϕ existed, we could introduce the composite $h \equiv G \circ \phi$, where again G is as above. Clearly, if x is rational, $\phi(x)$ is irrational, so G is continuous at $\phi(x)$ and thus h is continuous at x.

On the other hand, h will be discontinuous at any irrational y. To see this, choose $\{x_n\}$, a sequence of *rationals* converging to y. Then,

$$\lim_n h(x_n) = \lim_n G(\phi(x_n)) = \lim_n 0 = 0,$$

since $\phi(x_n)$ is irrational for each n and G is zero on the irrationals. But $h(y) = G(\phi(y)) \neq 0$ because $\phi(y)$ is a rational number.

In short, the function h defined above has C_h equal to the set of rationals, an impossibility by Volterra's earlier observation. Thus, it is impossible continuously to transform rationals to irrationals and vice versa.

With these reasonably elementary arguments—a tribute to the genius of Vito Volterra—we can use some of yesterday's mathematics to answer today's questions. The inquisitive student should be both satisfied and, one hopes, impressed. And in this instance the history of mathematics comes alive by proving its value in the contemporary classroom.

REFERENCES

1. R. Baire, Sur les fonctions de variables réelles, *Annali di Matematica pura ed applicata* 3 (1899), 1–122.
2. C. Clark, *The Theoretical Side of Calculus*, Krieger, 1978.

3. H. Hankel, Untersuchungen über die unendlich oft oszillierenden und unstetigen Functionen, presented in March, 1870 at the University of Tübingen; reprinted in *Mathematische Annalen* 20 (1882), 63–112.
4. T. Hawkins, *Lebesgue's Theory of Integration: Its Origins and Development*, Chelsea, 1975.
5. B. Riemann, *Gesammelte Mathematische Werke und Wissenschaftlicher Nachlass*, 2nd ed., Leipzig, B. G. Teubner, 1902.
6. H. Royden, *Real Analysis*, 2nd ed., Macmillan, 1968.
7. H. J. S. Smith, On the integration of discontinuous functions, *Proc. London Math Soc.* 6 (1875), 140–153.
8. V. Volterra, Alcune osservasioni sulle funzioni punteggiate discontinue, in *Opere Matematiche* 1 (Rome 1954), 7–15 (originally in *Giornale di Matematiche* 19 (1881), 76–86).

Einstein's Principle

B. L. FOSTER
University of Montana
Missoula, MT 59812

In his remarkable Spencer Lecture, delivered at Oxford in 1933, Albert Einstein advanced an astonishing epistemological view:

> If then it is the case that the axiomatic basis of theoretical physics cannot be an inference from experience, but must be free invention, have we any right to hope that we shall find the correct way?... To this I answer with complete assurance, that in my opinion there is *the* correct path and, moreover, that it is in our power to find it. Our experience up to date justifies us in feeling sure that in Nature is actualized the ideal of mathematical simplicity. It is my conviction that pure mathematical construction enables us to discover the concepts and the laws connecting them which give us the key to the understanding of the phenomena of Nature... In a certain sense, therefore, I hold it to be true that pure thought is competent to comprehend the real, as the ancients dreamed.
>
> To justify this confidence of mine, I must necessarily avail myself of mathematical concepts. The physical world is represented as a four-dimensional continuum. If in this I adopt a Riemannian metric, and look for the simplest laws which such a metric can satisfy, I arrive at the relativistic gravitation-theory of empty space. If I adopt in this space a vector-field,..., and if I look for the simplest laws which such a field can satisfy, I arrive at the Maxwell equations for free space. [**1**, p. 167]

The elevation of mathematical simplicity from helpful adjunct to indispensable guiding principle for the discovery of natural law seems radical even today. That so many of his last years were spent in vain on a quest for the simple field which would comprise, as special cases, gravity's Riemannian metric and electromagnetism's vector-field, may have unduly obscured the basic soundness of the principle, which he believed to be founded on the rock of general relativity.

This note gives an elementary illustration of the principle applied to pre-relativity physics. If a vector-field is thought of as first rank, and a metric as second rank, then the next step down is a scalar-field, of rank zero. For it the following, itself based on a hint of Einstein's (in [**2**, p. 29]), holds: