Historia Mathematica **28** (2001), 18–30
doi:10.1006/hmat.2000.2301, available online at http://www.idealibrary.com on IDEAL®

Newton and the Notion of Limit

Bruce Pourciau

Department of Mathematics, Lawrence University, Appleton, Wisconsin 54912

E-mail: bruce.h.pourciau@lawrence.edu

We investigate Newton's understanding of the limit concept through a study of certain proofs appearing in the *Principia*. We find that Newton, not Cauchy, was the first to present an epsilon argument, and that, in general, Newton's understanding of limits was clearer than commonly thought. We observe Newton's distinction between two properties easily confused, namely $f/g \to 1$ and $f - g \to 0$, we resolve a problem created by a spurious translation appearing in Cajori's revision of Motte's original translation, and we come to a deeper understanding of the well-known but less well understood Lemma XI of Section I, Book I. © 2001 Academic Press

Nous examinons la notion newtonienne du concept de limite en étudiant certaines preuves qui apparaissent dans les *Principia*. Nous découvrons que Newton, et non pas Cauchy, a été le premier à présenter un argument d'epsilon et que, en général, la compréhension newtonienne de la limite était bien plus lucide qu'on ne le pense communément. Nous observons la distinction que fait Newton entre deux propriétes qui se confondent facilement, à savoir $f/g \to 1$ et $f - g \to 0$, et nous résolvons un problème né d'une traduction incorrecte parue dans la révision par Cajori de la traduction de Motte. Nous parvenons ainsi à comprendre plus complètement le Lemma XI, Section I, Livre I, bien connu mais moins bien compris. © 2001 Academic Press

Key Words: Newton; *Principia*; limit; ultimate ratio; Cauchy.

How well did Newton understand the notion of limit? In chapters on the evolution of the limit concept in standard books on the history of mathematics, Newton comes across as vague and confused, almost clueless, especially compared, for example, to Cauchy:

Newton.

He had no approach to a limit that would be recognized today. [Bell 1945, 151]
This is the clearest statement Newton gave as to the nature of ultimate ratios, but ... it is precisely this lack of arithmetical clarity which led to controversial discussions ... as to what Newton really meant.... The meanings of the terms ... "prime and ultimate ratio" had not been clearly explained by Newton, his answers being equivalent to tautologies. ...Such an interpretation of Newton's meaning, which of course results in the ... indeterminate ratio 0/0, is not unjustified. [Boyer 1959, 198, 216, and 226]
Though by no means clear, this is the clearest statement Newton ever gave.... [Kline 1980, 135]
In this ... there are obscurities and difficulties. Newton appears to teach that a variable quantity and its limit will ultimately coincide and be equal. [Cajori 1991, 199]
Newton [spoke] of "ultimate values of vanishing quantities," but this is only to cover up with words the imprecision of the ideas. [Dieudonné 1992, 62]
[Newton] ... succeeded in doing [nothing] more with the limit concept than confusing himself ... [Kline 1953, 231]

Cauchy.

The definitions of limit ... current today in thoughtfully written texts on elementary calculus are substantially those expounded and applied by Cauchy. [Bell 1945, 292]

0315-0860/01 $35.00

> This is the most clear-cut definition of the concept which had been given up to that time. [Boyer 1959, 273]
>
> It is essentially [Cauchy's] definitions that we find in the more carefully written of today's elementary textbooks on the calculus. [Eves 1983, 425]
>
> Even today Cauchy's definition of limit ... will be found in any carefully written book on the calculus. [Bell 1937, 286]
>
> Cauchy's work ... presented an acceptable set of definitions ... needed for a rigorous calculus. [Fauvel and Gray 1987, 563]
>
> He defined carefully ... the basic notions of the calculus—function, limit.... [Kline 1980, 174]
>
> With the teaching work of Cauchy ... we find ourselves at last on solid ground. ...The notion of limit, fixed once and for all, is taken as the point of departure. [Bourbaki 1994, 196]

Based on these statements, a search through original sources for the actual definitions of limit given by these two mathematicians should reveal the stark contrast of a muddled Newton and a rigorous Cauchy. The leading section of Newton's *Principia* opens with eleven preliminary mathematical lemmas and closes with a discussion of limits, in particular limits of quotients, or, as Newton puts it, ultimate ratios, and it is here in this closing discussion that we find Newton's best definition of limit:

> Those ultimate ratios ... are not actually ratios of ultimate quantities, but limits ... which they can approach so closely that their difference is less than any given quantity.... [Newton 1999, 442; Newton 1946, 39][1]

A surprise: this is not the confused Newton we were led to expect. It may be more an epsilon than an epsilon-delta definition, but the core intuition is clear and correct. Well, perhaps Newton's definition will look more confused when we compare it to Cauchy's 1821 definition, which comes from his *Cours d'Analyse*:

> When the successively attributed values of the same variable indefinitely approach a fixed value, so that finally they differ from it by as little as desired, the last is called the *limit* of all the others. [Grabiner 1981, 80]

Surprised again: where is the rigorous Cauchy, the Cauchy whose "definition of limit ... will be found in any carefully written book on the calculus"? One would be hard pressed to squeeze 34 much less 134 (=1821–1687) years worth of difference between the definitions given by Newton and Cauchy. What's going on?

Partly the problem stems from our tight focus on the definitions, a narrowness that has inadvertently led us to compare the worst of Cauchy with the best of Newton and to obtain as a result a misleading gauge of their relative command of the limit notion. In modern mathematics, we naturally look for the meaning of a concept in its given definition, but in Newton's work and Cauchy's as well, the intended meaning often lies hidden in a relatively vague definition and only comes out explicitly and more precisely in subsequent calculations and proofs. For example, in the *Principia* the more precise meaning behind Newton's rough definition of acceleration—as a quantity "proportional to the velocity which it generates in a given time" [Newton 1999, 407; Newton 1946, 4]—only becomes clear later, in the demonstrations of certain following propositions, where Newton computes accelerations with a careful limit argument. Similarly, the real meaning of Cauchy's definition of limit becomes manifest in later proofs, where, given an , Cauchy determines an appropriate .

[1] Our translations throughout will follow Cohen and Whitman [Newton 1999], but we give page numbers for both [Newton 1999] and Cajori's revision [Newton 1946] of Andrew Motte's 1729 translation.

Concentrating on just his *definition*, we thus underestimate Cauchy's *understanding* of limits. Do we make the same error with Newton, but in the opposite direction? Singling out his definition from the Principia, do we *overestimate* Newton's understanding of limits? We shall argue here that the answer is no, at least with regard to Newton's *mature* understanding, that in fact by the time he wrote the *Principia* in 1687, Newton's understanding of the limit process was as good as his surprisingly good definition and significantly better than many works on the history of mathematics would have us believe. This position would be difficult to defend if Newton's definition stood alone. A single call of clarity never duplicated, never amplified or applied in any demonstration, could well seem almost accidental. But as we shall see, Newton *does* apply his definition while proving three preliminary mathematical lemmas of the *Principia*, and a study of these proofs will tell us more about Newton's understanding of limits (including the limits of that understanding) and will convince us, in particular, that the clarity of his definition was no accident.

As we noted earlier, the statement we have taken as Newton's definition of limit appears early in the *Principia* in a discussion of the limit concept that follows the eleven preliminary mathematical lemmas of Section I. Let us return to that point in the *Principia*, before we go on to the lemmas themselves, because the two sentences which follow the definition provide special insight into Newton's command of the limit notion. At this stage of the discussion, Newton tries to make clear (without the benefit of function notation) that the limit of a ratio f/g, where both f and g tend to 0, is not the ratio of the limits:

> It can also be contended, that if the ultimate ratios of vanishing quantities [that is, the limits of such ratios] are given, their ultimate magnitudes will also be given; and thus every quantity will consist of indivisibles, contrary to what Euclid has proved.... But this objection is based on a false hypothesis. Those ultimate ratios with which quantities vanish are not actually ratios of ultimate quantities, but limits which ... they can approach so closely that their difference is less than any given quantity... *This matter will be understood more clearly in the case of quantities indefinitely great. If two quantities whose difference is given are increased indefinitely, their ultimate ratio will be given, namely the ratio of equality, and yet the ultimate or maximal quantities of which this is the ratio will not on this account be given*. [Newton 1999, 443; Newton 1946, 39, italics added]

In the calculus classroom, students always want to "plug in," to calculate limits, whether it's plugging in the limits of the numerator and denominator to compute the limit of the ratio or plugging in the value of the function to find the limit of the function. To get them to see, not just the possibility of error in this, but the *misunderstanding* of the limit notion it implies, a good teacher might provide his class with many examples where one manifestly cannot plug in at all. One such example: the limit of a ratio f/g where the difference $f - g$ remains fixed as both f and g become arbitrarily large. Newton, a good teacher often in the *Principia* (but not in the lecture hall, where he cared little and students responded in kind), offers exactly this example in the two sentences above that we have stressed with italic. As he correctly points out, in a case like this $f/g \to 1$ even though f and g themselves have no limits at all. (A simple instance of this: $(t+1)/t \to 1$ as $t \to +\infty$.) Newton's illustration is the excellent pedagogical choice of a teacher who knows his subject well, who knows in particular the distinction between having a ratio that tends to 1 and having a difference that tends to 0, that one can have, as in his example, $f/g \to 1$ without $f - g \to 0$.

Ironically, just at this point where Newton, in making this distinction, illustrates his grasp of the limit concept, a 19th-century commentator on the *Principia*, J.F.M. Wright, illustrates his own confusion:

> The fact is, Newton himself, if we may judge from his own words..., "If two quantities, whose difference is given, be augmented continually, the ultimate ratio of these quantities will be a ratio of equality," had no knowledge of the true nature of his method of prime and ultimate ratios. If there be meaning in words, he plainly supposes in this passage, a mere approximation to be the same with an ultimate ratio. [Wright 1833,3]

Thus Wright, in 1833, believed that $f/g \to 1$ necessarily implies $f - g \to 0$, but Newton, in 1687, knew better. Not to dwell on the confusion of Principian commentators, since it does little to extend our case for Newton's understanding of limits, but Wright also believed the converse, namely that $f - g \to 0$ entails $f/g \to 1$, which is false as well: $t - t^2 \to 0$ if $t \to 0$ with $t > 0$, yet t/t^2 becomes arbitrarily large. Our study of the preliminary mathematical lemmas and their proofs, even just the three we take up below, will reveal that Newton knew exactly when these conditions, $f/g \to 1$ and $f - g \to 0$, were equivalent.

Let us then turn to these lemmas that fill out Section I, Book One, of the *Principia*. "I have presented these lemmas before the propositions," Newton tells us,

> to avoid the tedium of working out lengthy proofs by *reductio ad absurdum*, in the manner of the ancient geometers. ...I preferred to make the proofs of what follows depend on the ultimate sums and ratios of vanishing quantities [instead of the method of indivisibles]... For the same result is obtained by these as by the method of indivisibles, and we shall be on safer ground using principles that have been proved. [Newton 1999, 441; Newton 1946, 38]

When he composed the lemmas, Newton may well have thought of them as foundational, as basic results in his geometric calculus that he would refer to frequently in subsequent arguments of the *Principia*. On the other hand, he may have inserted them later, creating a sort of "retrospective gloss" as Whiteside puts it [Newton 1967–1981 VI, 108], to increase the apparent rigor of the exposition. Whatever the case, these eleven preliminary mathematical lemmas are not in fact cited often in the demonstrations which follow Section I, but they are, nonetheless, truly foundational in a different sense, for not only is each lemma the geometric version of a basic definition, property, or theorem of calculus, but taken together the lemmas form a coherent and natural whole: in their geometric disguise, they are just the calculus results one would expect to need in any mathematical study of orbital motion. More on this aspect of the lemmas can be found in [Pourciau 1998].

In the present study, we have a particular interest in three of these lemmas—Lemma I on the limit of a difference, Lemma II on the existence of the integral, and Lemma XI on the second derivative—because their statements and proofs most clearly reveal Newton's grasp of the limit process. To read these lemmas requires a double translation, not only a first translation from the original Latin into English (for which we rely on [Newton 1999]), but then a second translation as well, for the lemmas come to us packed in the *Principia*'s unique blend of Euclidean geometry and limits, a sort of geometric calculus, and we cannot sort out what the lemmas really say without doing some unpacking. But any translation disturbs meaning, and we must take great care to minimize that disturbance, to preserve as far as possible Newton's original intent.

We begin with Lemma I and its brief demonstration:

> LEMMA I. *Quantities, and also ratios of quantities, which in any finite time constantly tend to equality, and which before the end of that time approach so close to one another that their difference is less than any given quantity, become ultimately equal.*
>
> If you deny this, let them become ultimately unequal, and let their ultimate difference be D. Then they cannot approach so close to equality that their difference is less than the given difference D, contrary to the hypothesis. [Newton 1999, 433; Newton 1946, 29]

Let us try to tease the meaning out line by line. To introduce "quantities, and also ratios of quantities" is to introduce, without the benefit of function notation, two functions, f and g say, which themselves may be ratios of other functions. To suppose of f and g that "before the end of [some finite] time [they] approach so close to one another that their difference is less than any given quantity," is to suppose that as $t \to c^-$ (that is, as t approaches c from the *left*), where c is finite and positive, $f(t) - g(t) \to 0$. When Newton requires that f and g "*constantly* tend to equality," [my italics], he is assuming further that $f(t) - g(t)$ not only converges to 0, but actually decreases *monotonically* to 0. (We may abbreviate this by writing $f(t) - g(t) \downarrow 0$.) Under these conditions, the lemma concludes that the limits $l \equiv \lim_{t\to c^-} f(t)$ and $m \equiv \lim_{t\to c^-} g(t)$ exist and are equal.

Newton assumes c finite presumably to rule out cases like $f(t) = t, g(t) = t + (1/t)$, where $f - g \downarrow 0$ as $t \to +\infty$, yet the individual limits of f and g fail to exist. Of course taking c finite does not exclude an example such as $f(t) = 1/(1-t) + 1 - t, g(t) = 1/(1-t)$, where again $f - g \downarrow 0$ (as $t \to 1^-$) while f and g themselves grow arbitrarily large, but then part of the meaning of Newton's "quantities" may have been that they could not become arbitrarily large in finite time. Moreover, these "quantities" (which are generally lengths or areas of geometric figures) appear by implicit assumption to be smooth (in fact to possess as many derivatives as required) and (at least locally) monotonic.

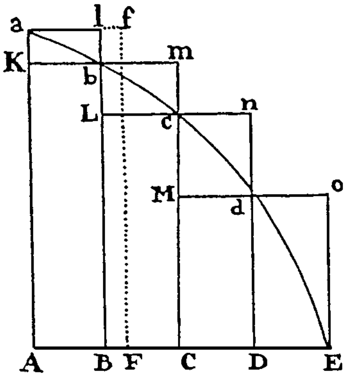
With these constraints, the ultimate values of f and g (that is, $l \equiv \lim_{t\to c^-} f(t)$ and $m \equiv \lim_{t\to c^-} g(t)$) must indeed exist, and the proof of Lemma I is simple and correct. For if we deny the conclusion, $D \equiv l - m$ would be positive, not zero, and then the inequality $D < f(t) - g(t)$ (for $t < c$), which follows from the assumption in the lemma that the difference $f(t) - g(t)$ decreases monotonically, would prevent that difference from becoming as small as we please, a contradiction.

Certain aspects of Lemma I remain unclear, and our interpretation of Newton's intended meaning must be somewhat tentative. For example, is it reasonable (that is, how much distortion does it cause) to think of Newton's "quantities" as functions? Continuous functions? Smooth functions? Is the existence of the limits l and m an implicit hypothesis of this particular lemma, a consequence of some background assumption about "quantities," a result of assumptions in the lemma, presumed in error, or the product of some combination of these possibilities? What *does* remain clear, however, is Newton's core understanding, here a special case of his definition of limit, that a quantity has limit zero if it can be made less than any given distance. Simple and clear. No indivisibles nor "ghosts of departed quantities." No confusion.

The same basic clarity comes out in Lemma II and its proof:

> LEMMA II. *If in any figure $AacE$, comprehended by the straight lines Aa and AE and the curve acE, any number of parallelograms Ab, Bc, Cd, ... are inscribed upon equal bases AB, BC, CD, ... and*

sides, Bb, Cc, Dd, ... parallel to the side Aa of the figure; and if the parallelograms $aKbl$, $bLcm$, $cMdn$, ... are completed; if then the width of these parallelograms is diminished and their number increased indefinitely, I say that the ultimate ratios which the inscribed figure $AKbLcMdD$, the circumscribed figure $AalbmcndoE$, and the curvilinear figure $AabcdE$, have to one another are ratios of equality.



For the difference of the inscribed and circumscribed figures is the sum of the parallelograms Kl, Lm, Mn, and Do, that is (because they all have equal bases), the rectangle having as base Kb (the base of one of them) and as altitude Aa (the sum of the altitudes), that is, the rectangle $ABla$. But this rectangle, because its width AB is diminished indefinitely, becomes less than any given rectangle. Therefore, (by lem. I) the inscribed figure and the circumscribed figure and, all the more, the intermediate curvilinear figure become ultimately equal. Q.E.D. [Newton 1999, 433; Newton 1946, 29]

In notation more comfortable for us (and doing no real damage to Newton's original meaning), the figure $AacE$ would be the graph of a function f defined on the segment AE. Newton clearly (but without saying so) takes f to be monotone decreasing with $f(E) = 0$. Of course the areas of the inscribed and circumscribed figures, $AKbLcMdD$ and $AalbmcndoE$, correspond to lower and upper sums,

$$L_n \equiv f(t_1)\Delta t + \cdots + f(t_n)\Delta t \quad \text{and} \quad U_n \equiv f(t_0)\Delta t + \cdots + f(t_{n-1})\Delta t,$$

that arise from a partition

$$A = t_0 < t_1 < \cdots < t_{n-1} < t_n = E$$

of the segment AE into n subintervals of equal length $\Delta t = AE/n$. If by $\mathcal{A}$ we denote the area under the graph of f (an area which Newton takes here to be computed, not defined), then the lemma concludes: "the ultimate ratios [which L_n, $\mathcal{A}$, and U_n] have to one another are ratios of equality," that is, the ratio of any one to the other will tend to 1 as $n \to \infty$.

Turning to Newton's proof, he notes that the difference of the inscribed and circumscribed areas is less than the area of the rectangle $ABla$; that is,

$$U_n - L_n < Aa \cdot AB = Aa \cdot \frac{AE}{n}.$$

Then he remarks that the area of this rectangle "becomes less than any given rectangle," and concludes, from Lemma I, that

$$\lim_{n\to\infty} L_n = \mathcal{A} = \lim_{n\to\infty} U_n.$$

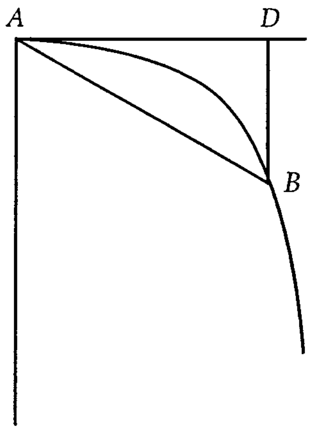
In his geometric style, Newton has stated and proved a basic theorem of calculus—

THEOREM *Every monotonic function on a closed and bounded interval must be integrable.*

—and his demonstration is essentially the proof "that we find in the more carefully written of today's elementary textbooks on the calculus," to borrow from [Eves 1983, 425] an appreciation of Cauchy. We also see again the fundamental understanding and definition of a limit as a value approached "so closely that their difference is less than any given quantity." [Newton 1999, 442; Newton 1946, 39] On the other hand, having earlier seen Newton give an example where $f/g \to 1$ yet $f - g$ does not go to 0, we now find him apparently equating $L_n/U_n \to 1$ ("ratios of equality," in the statement of the lemma) with $U_n - L_n \to 0$ ("difference ... becomes less than any given rectangle," in his demonstration). Has he forgotten himself? Surely not. He just knows the conditions under which these two limit properties are equivalent—namely, that either the limit of f or the limit of g is finite and nonzero.

Impressed again with Newton's understanding of the limit process, let us skip ahead to the somewhat more obscure yet enlightening case of Lemma XI. All along our translations have followed Cohen and Whitman [Newton 1999], but for reasons that will become clear soon, in the case of Lemma XI we first give the version found in Cajori's revision of Motte's original 1729 translation:

LEMMA XI. *The evanescent subtense of the angle of contact, in all curves which at that point of contact have a finite curvature, is ultimately as the square of the subtense of the conterminous arc.* [Newton 1946, 36]



Some translation work of our own would seem to be in order here, since the English sounds as foreign as the original Latin. In the figure above, the line AD is tangent to the curve AB at the point A. By the "evanescent subtense of the angle of contact" Newton means the length BD and by the "subtense of the conterminous arc" he means the length AB of the chord. (For simplicity and to follow the *Principia*'s style, AB and BD, and similar notations, may refer to curves, lines, segments, or segment lengths. Context and verbal cues—for instance, "the length AB"—should help us from getting confused.)

Furthermore, one quantity "is ultimately as" another, according to Newton, when their ratio tends to a finite positive number. Thus to claim that the "evanescent subtense of the angle of contact ... is ultimately as the square of the subtense of the conterminous arc" boils down to claiming that the ratio BD/AB^2 tends to a finite positive number as D tends to A. Intuitively, this tells us that in a small neighborhood of the point A, BD is just about proportional to AB^2, since in such a neighborhood their ratio remains almost (a positive) constant.

Now the lemma makes this claim, not for all curves, but only those with "finite curvature" at A. Note first that Newton's original Latin, *finitam*, translated naturally as "finite," should actually be taken as meaning "finite and nonzero" or perhaps "finite and positive" in many contexts, including this one, in part because a vanished quantity in the *Principia* (often a length, area, or volume) is not considered "finite," like the number zero, but gone, like a nonexistent length. More importantly, when the assumption "...finitam" appears in a lemma or proposition, we can often tell from the way Newton applies this assumption in a subsequent demonstration that "finite and nonzero" is the intended meaning.

As for "curvature," this needs no translation, for it means to us what it did to Newton. Just to remind ourselves, the curvature at any point of a circle is the reciprocal of the radius, and the curvature at the point A of a curve AB is the curvature of that circle which "fits the curve best" at A. (We have no need for the precise definition here.) As early as 1671, in his long tract on series and fluxions [Newton 1967–1981 III, 32–353], Newton had defined curvature for plane curves, given several equivalent characterizations, obtained formulas for it, one involving partial derivatives, and calculated the curvature of many different curves. A comment in his *Waste Book*, recorded in December 1664 or January 1665, indicates that he also understood the connection between curvature and the mathematical study of orbital motion: "If the body b moved in an Ellipsis, then its force in each point (if its motion in that point bee given) may bee found by a tangent circle of equall crookednesse with that point of the Ellipsis" [Whiteside 1991, 14]. In spite of this, curvature remains generally behind the scenes in the first (1687) edition of the *Principia*. By 1694, however, as he planned for the second edition, Newton had drafted extensive revisions for the early sections of Book One following a scheme that actually *centered* on curvature. (See [Brackenridge 1992; Newton 1967–1981 VI; Pourciau 1992) In the end, these radical revisions never made it into the *Principia*, and Newton had to content himself with pasting little pieces of his grand scheme into the second (1713) edition as added assumptions, brief asides, new corollaries, and alternate demonstrations.

Our Lemma XI carries a reminder of this renewed attention to curvature. In the second edition (in fact in his annotated copy of the first edition), Newton added the assumption that the arc AB has finite [and nonzero] curvature at the point A [Newton 1972, 83]. And with this assumption, we can see (from a modern perspective) why the conclusion of Lemma XI (as translated in Cajori's revision [Newton 1946] of Motte) should hold. If we think of the arc AB as the graph of a smooth function f with $f(0) = 0 = f'(0)$, then the standard calculus formula for the curvature at $t = 0$ gives

$$(0) = \frac{f''(0)}{[1 + f'(0)^2]^{3/2}} = f''(0).$$

On the other hand, for x near 0, since f is smooth,

$$\begin{aligned} f(x) &= f(0) + f'(0)x + \frac{1}{2!} f''(0)x^2 + \frac{1}{3!} f'''(0)x^3 + \cdots \\ &= \frac{1}{2!} f''(0)x^2 + \text{terms involving } x^3 \end{aligned}$$

and thus

$$\frac{f(x)}{x^2} \to \frac{1}{2} f''(0) \quad \text{as } x \to 0.$$

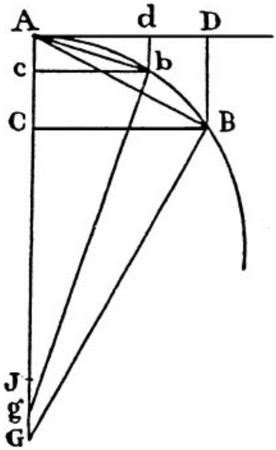
Given that $\;(0) = f''(0)$ is finite and positive, we could therefore use Newton's phrasing and say that $f(x)$ is ultimately as x^2, or, using his notation as well, that BD is ultimately as AD^2. But since $AD/AB \to 1$ (from the earlier Lemma VII or from the fact that AD/AB is the cosine of the vanishing angle), we see that BD is ultimately as AB^2, which is the conclusion of Lemma XI (in [Newton 1946]).

Naturally the demonstration of Lemma XI in the *Principia* has a more geometric flavor, for Newton cultivated classical geometry in the *Principia* generally, hoping to inherit the clarity, simplicity, elegance, and rigor he saw in classical geometric reasoning. But Newton's actual proof of Lemma XI (which we take up soon), though it contains a lovely geometric kernel, seems at first both less elegant and less rigorous than we might expect—less elegant because apparently extraneous points and lines are introduced and less rigorous because the conclusion of the proof does not quite match the conclusion of the lemma (as given in [Newton 1946]). Moreover, after we read Newton's demonstration, an obvious revision immediately presents itself, a simpler argument based on the same geometric construction, and one naturally wonders why this simpler, correct argument fails to appear in the *Principia*.

Here is that simpler argument: Draw BG perpendicular to the line AB. (See Newton's original figure for Lemma XI below. It contains a couple of points and lines we do not require at the moment.) As D approaches A, the point G tends to J (that is, the length GJ can be made less than any assignable distance), where AJ is a finite, nonzero length. [That AJ is a finite, nonzero length follows directly from Newton's definition of curvature, which, ironically, the *Principia* omits. In fact, AJ turns out to be the diameter of curvature at A.] By the nature of the circle passing through A, B, and G, we have $AB^2 = AG \cdot BD$ (because the triangles ABG and ABC are similar) and it follows that $AB^2/BD = AG$ is ultimately AJ. Hence BD is ultimately as $AB^2 \cdot$ QED.

An elegant and simple geometric proof. Why doesn't Newton give it? In part because, following Euclidean conventions, Newton prefers to regard his ratios, not as quotients of numbers, but as proportions between geometric quantities, indeed as proportions between geometric quantities *of the same kind*: lengths to lengths, areas to areas, volumes to volumes. This leads him to introduce points d, b, and g, in addition to D, B, and G, so that instead of comparing an area AB^2 to a length BD, he can compare areas to areas (AB^2 to Ab^2) and lengths to lengths (BD to bd).

He then has D and d approach A together, in the sense that Ad is ultimately as AD (at least this is my interpretation of his directions) and proves that ultimately $AB^2/Ab^2 = BD/bd$, that is, that $\lim_{D\to A} AB^2/Ab^2 = \lim_{D\to A} BD/bd$.



Here then we see an (uncommon) instance where Newton's desire to be Euclidean (rather than just geometric) has led to a *less* elegant construction. But more interestingly, and more potentially damaging to our confidence in Newton's understanding of limits, notice the apparent error in the argument, for what has been proved here is not actually *quod erat demonstrandum* (at least not according to the translation in [Newton 1946]), but something else. Newton does prove that $\lim_{D\to A}(BD/bd)$ is finite, nonzero, and equal to $\lim_{D\to A}(AB^2/Ab^2)$, but from this it does not automatically follow that $\lim_{D\to A}(AB^2/BD)$ is finite and nonzero, that is, that BD is ultimately as AB^2, which is (in [Newton 1946]) *quod erat demonstrandum*. For example,

$$\lim_{t\to 0^+}\frac{t}{t} = 1 = \lim_{t\to 0^+}\frac{t^2}{t^2}, \quad \text{and} \quad \text{yet} \quad \lim_{t\to 0^+}\frac{t}{t^2} = +\infty = \lim_{t\to 0^+}\frac{t}{t^2}.$$

Has his need to be Euclidean, in this place at least, led Newton not only to needless complication, but to error as well?

No. In fact the error lies not with Newton but with a spurious translation of Lemma XI in Cajori's revision [Newton 1946] of Motte's translation. Here is the original Latin from the third (1726) edition of the *Principia*, as recorded in [Newton 1972, 83]:

> *Subtensa evanescens anguli contactus, in curvis omnibus curvaturam finitam ad punctum contactus habentibus, est ultimo in ratione duplicata subtensæ arcus contermini.*

Motte translated this correctly in 1729 and Cohen and Whitman have translated it correctly in 1999:

> In all curves having a finite curvature at the point of contact, the vanishing subtense of the angle of contact is ultimately in the squared ratio of the subtense of the conterminous arc. [Newton 1999, 439]

To claim that the "vanishing subtense of the angle of contact is ultimately in the squared ratio of the subtense of the conterminous arc," is to claim that $\lim_{D\to A} AB^2/Ab^2 = \lim_{D\to A} BD/bd$, and this is precisely what Newton proves. Cajori—or perhaps more likely R.T. Crawford, who edited Cajori's manuscripts after Cajori's death in 1930—striving to modernize the Motte translation, changes “ultimately in the squared ratio of the subtense of the conterminous arc” to “ultimately as the square of the subtense of the conterminous arc,” and while this revised conclusion to Lemma XI is indeed more modern, and still true, it is not what Newton asserts and it is not what Newton proves. (See [Newton 1999, 26–42] for a discussion of the difficulties involved in translating the *Principia*.)

Thus the *Principia*’s demonstration of Lemma XI is in fact correct. Earlier an outline of that demonstration was enough, but now we need to study the details to see what they tell us about Newton’s grasp of the limit concept. We omit some preliminary lines where the notation (referring to the figure above) is introduced:

> ...let J be the intersection of lines BG and AG, which ultimately occurs when points D and B reach A. It is evident that the distance GJ can be less than any assigned distance. And (from the nature of the circles passing through points A, B, G and a, b, g) AB^2 is equal to $AG \times BD$, and Ab^2 is equal to $Ag \times bd$, and thus the ratio of AB^2 to Ab^2 is compounded of the ratios of AG to Ag and BD to bd. *But since GJ can be taken as less than any assigned length, it can happen that the ratio of AG to Ag differs from the ratio of equality by less than any assigned difference, and thus that the ratio of AB^2 to Ab^2 differs from the ratio of BD to bd by any assigned difference. Therefore, by lem. 1, the ultimate ratio of AB^2 to Ab^2 is the same as the ultimate ratio of BD to bd*. Q.E.D. [Newton 1999, 439, italics added]

If we look at the lines in italic, we see what may be the first algebraic epsilon-argument ever given! It is simple, but correct, and it shows that by 1687 Newton had acquired a surprisingly clear conception of the limit process. We remain impressed, here in this proof and in the early sections of the *Principia* generally, by Newton’s mastery of the basic idea. There are no infinitely small quantities, no ratios of indivisibles, no plugging-in to compute the limit, no kinematic comments about velocity. Just the correct fundamental question: Can we make the difference between this quantity and this fixed value less than any given positive number?

Of course Cauchy would have been altogether more explicit. For instance, instead of merely claiming that the “ratio of AB^2 to Ab^2 differs from the ratio of BD to bd by any assigned difference,” Cauchy, in a clarifying function notation, would have given himself an , written

$$\frac{AB^2}{Ab^2} - \frac{BD}{bd} = \left(1 - \frac{Ag}{AG}\right)\frac{AB^2}{Ab^2},$$

noted that AB^2/Ab^2 was bounded (because AD is ultimately as Ad, which implies AB is ultimately as Ab), and used the algebra of inequalities to compute an appropriate .

Certainly then, Newton’s argument is not Cauchy’s, but still it is unexpectedly sophisticated, especially given the received opinion that Newton was more confused than clear on the notion of limits. Grabiner’s *The Origins of Cauchy’s Rigorous Calculus* probably has the most accurate appraisal of Newton’s understanding of the limit concept, but even the thoughtful and scholarly opinion received from this work fails to account for what we see in the proof of Lemma XI. Pointing to Newton’s definition of limit which follows the preliminary lemmas, Grabiner remarks on his “influential statements about the limit concept, in words that were to recur throughout the eighteenth century,” and yet she sees the

> history of the limit concept until 1810 [as] the gradual solution of the verbal problems implicit in Newton’s explanation: the eventual substitution of algebraic language for Newton’s kinematic expressions; the

> broadening of the limit concept to include variables that oscillate about their limits; and—crucially—the abandonment of concern over whether a variable reaches its limit.

A wise comment, but after our review of Lemmas I, II, and XI it becomes hard to see Newton as merely a confusion to be clarified. Indeed Newton's epsilon argument in Lemma XI comes within epsilon of epsilon arguments produced by D'Alembert in 1789 and Lacroix in 1802, and while Grabiner praises the work of D'Alembert and Lacroix for "freeing the limit concept from physics and making it algebraic," [Grabiner 1981, 83] she fails to note the similar work of Newton (in the proof of Lemma XI) 100 years earlier.

Of course at different times, in other places, and even in Section I (Book One) of the *Principia*, Newton gave us arguments and explanations less clear than those we have studied here. In fact, in his discussion following Lemma XI, but before he gives his definition of limit, Newton provides the following intuitive view of the limit concept:

> It may be objected that there is no such thing as an ultimate proportion of vanishing quantities, in as much as before vanishing the proportion is not ultimate, and after vanishing it does not exist at all. But by the same argument it could equally be contended that there is no ultimate velocity of a body reaching a certain place at which the motion ceases; for before the body arrives at this place, the velocity is not the ultimate velocity, and when it arrives there, there is no velocity at all. But the answer is easy: to understand the ultimate velocity as that with which a body is moving, neither before it arrives at its ultimate place and the motion ceases, nor after it has arrived there, but at the very instant when it arrives, that is, the very velocity with which the body arrives at its ultimate place and with which the motion ceases. And similarly the ultimate ratio of vanishing quantities is to be understood not as the ratio of quantities before they vanish or after they have vanished, but the ratio with which they vanish. [Newton 1999, 442; Newton 1946, 39]

However, such informal and kinematic descriptions should be seen, not as a confused definition, but as an attempt to give the reader some intuitive insight into the limit process. How can we be sure? Look at the proofs. These kinematic descriptions may have been the source for Newton's *intuition* about limits, but if they were much more, if they were in fact part of his *definition*, then these kinematic notions would appear in the proofs we have studied. But they do not. It is in the proofs that we see the intended mathematical meaning. To see how Cauchy understood the limit concept, we look at his proofs; to see how Newton understood the limit concept, we have done the same. For a measure of his mature understanding, we took the *Principia*, which was revised almost to the day he died, and examined proofs in the section he devoted specifically to the limit process. Our study shows that by 1687, 100 years before D'Alembert and 134 years before Cauchy, Newton had a very clear grasp of the limit concept and was far less confused than the most common portrayals have led us to believe. In [Grabiner 1983] Grabiner asks, "Who gave you the epsilon?" and answers Cauchy. Our work here suggests a different Q & A: Who found the first delta? Cauchy. Who gave us the first epsilon? Newton.

ACKNOWLEDGMENT

I thank J. Bruce Brackenridge for his helpful comments.

REFERENCES

Bell, E. T. 1937. *Men of mathematics.* New York: Simon & Schuster.

Bell, E. T. 1945. *The development of mathematics*. New York: McGraw–Hill.

Bourbaki, N. 1994. *Elements of the history of mathematics*. Berlin: Springer-Verlag.

Boyer, C. B. 1959. *The history of the calculus and its conceptual development*. New York: Dover.

Brackenridge, J. B. 1992. The critical role of curvature in Newton's developing dynamics. In *An investigation of difficult things: Essays on Newton and the history of exact sciences*, pp. 231–260. Cambridge, U.K.: Cambridge Univ. Press.

Cajori, F. 1991. *A history of mathematics*. New York: Chelsea.

Dieudonné, J. 1992. *Mathematics—The music of reason*. Berlin: Springer-Verlag.

Eves, H. 1983. *An introduction to the history of mathematics*. New York: Saunders.

Fauvel, J., & Gray, J. 1987. *The history of mathematics: A reader*. London: Macmillan.

Grabiner, J. V. 1981. *The origins of Cauchy's rigorous calculus*, Cambridge, MA: MIT Press.

Grabiner, J. V. 1983. Who gave you the epsilon? Cauchy and the origins of rigorous calculus. *The American Mathematical Monthly* **90**, 185–194.

Kline, M. 1953. *Mathematics in Western culture*. New York: Oxford Univ. Press.

Kline, M. 1980. *Mathematics: The loss of certainty*. New York: Oxford Univ. Press.

Newton, I. 1972. *Isaac Newton's Philosophiae naturalis principia mathematica, the third edition with variant readings*, A. Koyré & B. Cohen, Eds., with the assistance of A. Whitman, Vol. I. Cambridge, MA: Harvard Univ. Press.

Newton, I. 1946. *Sir Isaac Newton's mathematical principles of natural philosophy and his system of the world*, a revision by F. Cajori of A. Motte's 1729 translation. Berkeley: Univ. of California Press.

Newton, I. 1967–1981. *The mathematical papers of Isaac Newton*, D. T. Whiteside, Ed., 8 vols. Cambridge, UK: Cambridge Univ. Press.

Newton, I. 1999. *The Principia: Mathematical principles of natural philosophy*, translated by I. B. Cohen & A. Whitman, preceded by *A guide to Newton's Principia* by I. B. Cohen. Berkeley: Univ. of California Press.

Pourciau, B. 1992. Radical *Principia, Archive for History of Exact Sciences* **44**, 331–363.

Pourciau, B. 1998. The preliminary mathematical lemmas of Newton's *Principia, Archive for History of Exact Sciences* **52**, 279–295.

Whiteside, D. T. 1991. The prehistory of the *Principia* from 1664 to 1686. *Notes and Records of the Royal Society of London* **45**, 11–61.

Wright, J. M. F. 1972. *A commentary on Newton's* Principia: *Designed for the use of students at universities*, Vol. 1, facsimile reprint of the 1833 edition, with an introduction and supplement by I. B. Cohen. New York: Johnson Reprint Corporation.