Seldin, Jonathan P., “From Exhaustion to Modern Limit Theory”
Reprinted from *Proceedings of the Sixteenth Annual Meeting of the Canadian Society for History and Philosophy of Mathematics*, University of Victoria, 1990.

From Exhaustion to Modern Limit Theory

Jonathan P. Seldin*
Department of Mathematics
Concordia University
Montréal, Québec
Canada
seldin@antares.concordia.ca

All of us who have had to teach the modern ε-δ theory of limits to calculus students are aware of how difficult the students find it. Even students who do not study it in calculus, but postpone it until their first course in analysis find it extremely difficult.

Now the ε-δ theory of limits of functions is very closely related to the ε-*N* theory of limits of sequences; in fact, Landau (1960), a book whose aim is to teach beginning calculus with a completely rigorous theory, starts with limits of sequences and only then goes on to limits of functions. But this theory of limits of sequences seems almost as difficult for students as the theory of limits of functions.

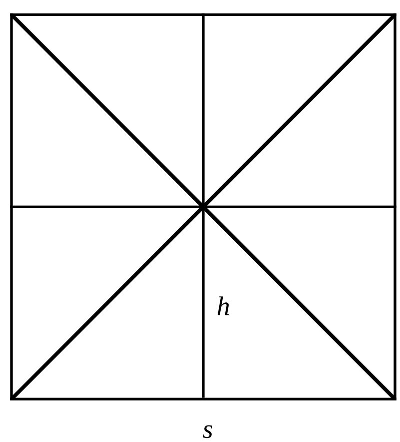
There is an older theory of what amounts to limits of sequences that goes back to Euclid and Archimedes. It is called the *method of*

* The work on which this document is based was supported in part by grant EQ41840 from the program Fonds pour la Formation de Chercheurs et l’aide à la Recherche (F.C.A.R.) of the Québec Ministry of Education and in part by grant OGP0023391 from the Natural Sciences and Engineering Research Council of Canada.

exhaustion, and it has the advantage that because it involves geometric figures it is possible for students to visualize what is going on. It is the purpose of this paper to show how it is possible to begin with proofs using the method of exhaustion in Archimedes and Euclid and, by writing out the steps in modern algebraic notation and doing some very simple manipulations arrive at modern ε-N proofs. I am presenting this as a proposal for introducing the theory of limits to students for the first time.

Judy Grabiner's paper elsewhere in this volume presents evidence that this approach approximates the historical route from Newton and Leibniz to Cauchy.

Let us begin with the problem of finding a formula for the area of a circle. This area is clearly closely related to the problem of finding the area of a regular polygon, so let us look at two examples of that problem. Let us begin with a square:



Consider each of the four triangles consisting of a side and the lines which join its ends to the centre. The area of each of these triangles is

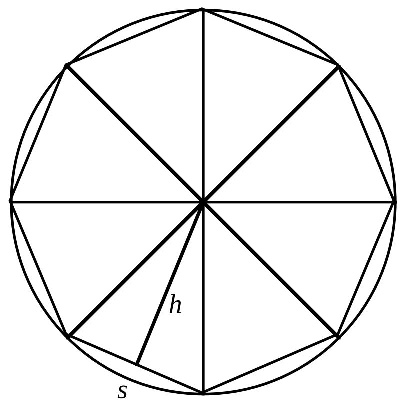
$$\frac{1}{2}hs.$$

The area of the entire square is the sum of the areas of the four triangles, namely,

$$\frac{1}{2}h(4s) = \frac{1}{2}hp,$$

where $p = 4s$ is the perimeter.

Now let us consider an octogon:



This time, we have eight triangles. As before, the area of each triangle is $(1/2)hs$, so the area of the octogon is

$$\frac{1}{2}h(8s) = \frac{1}{2}hp,$$

where this time the perimiter is $p = 8s$.

By now it should be clear that the area of *any* regular polygon is

$$\frac{1}{2}hp,$$

where h is what is called the “small radius”; i.e., the perpendicular from the center to a side, and where p is the perimeter. And since a circle appears to be essentially a regular polygon with an infinite number of sides, this suggests that the formula for the area of a circle is

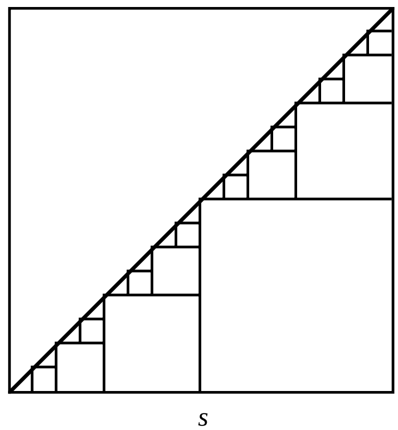
$$A = \frac{1}{2}rC.$$

It is not hard to see that this formula is correct by checking it with the formula that we are used to. Since $C = 2\pi r$, we have

$$A = \pi r^2 = \frac{1}{2}r(2\pi r) = \frac{1}{2}rC.$$

But as a method of establishing this formula, this is a circular procedure; indeed, we might call it a “proof by hindsight”.

The formula also *seems* obvious from the diagrams. But diagrams can be deceiving. For consider the following example:



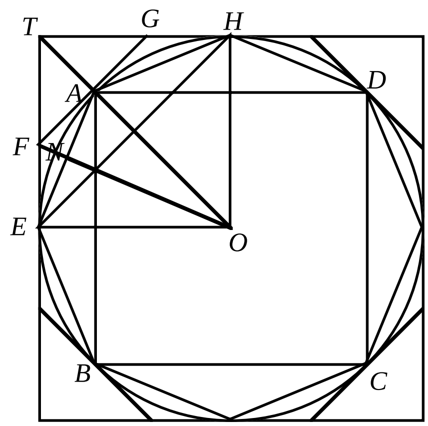
Now the length of the steped line is clearly $2s$, no matter how many steps there are. Furthermore, as the number of steps increases, the stepped line seems to approach the diagonal. But the length of the diagonal is $\sqrt{2}\, s \neq 2s$. This raises the question, *how can we tell when a limit which seems obvious from a diagram actually holds?*

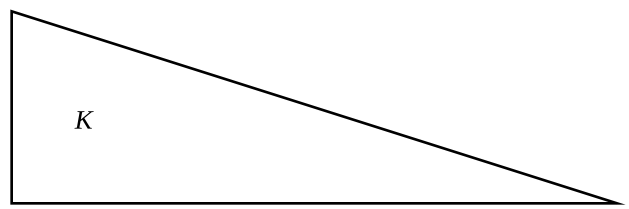
This problem was solved in ancient times, and we can read the solution in the works of Euclid and Archimedes. In the case of the area of a circle, the solution is due to Archimedes, and is found in his book "Measurement of a Circle" in Heath (1912):

Proposition 1.

> *The area of any circle is equal to a right-angled triangle in which one of the sides about the right angle is equal to the radius, and the other to the circumference, of the circle.*

Let $ABCD$ be the given circle, K the triangle described.





Then, if the circle is not equal to *K*, it must be either greater or less.

I. If possible, let the circle be greater than *K*.

Inscribe a square *ABCD*, bisect the arcs *AB*, *BC*, *CD*, *DA*, then bisect (if necessary) the halves, and so on, until the sides of the inscribed polygon whose angular points are the points of division subtend segments whose sum is less than the excess of the circle over *K*.

Thus the area of the polygon is greater than *K*.

Let *AE* be any side of it, and *ON* the perpendicular on *AE* from the center *O*.

Then *ON* is less than the radius of the circle and therefore less than one of the sides about the right angle in *K*. Also, the perimeter of the polygon is less than the circumference of the circle, i.e. less than the other side about the right angle in *K*. Also, the perimeter of the polygon is less than the circumference of the circle, i.e. less than the other side about the right angle in *K*.

Therefore the area of the polygon is less than *K*; which is inconsistent with the hypothesis.

Thus the area of the circle is not greater than *K*.

II. If possible, let the circle be less than K.

Circumscribe a square, and let two adjacent sides, touching the circle in E, H, meet in T. Bisect the arcs between adjacent points of contact and draw the tangents at the points of bisection. Let A be the middle point of the arc EH, and FAG the tangent at A.

Then the angle TAG is a right angle.

Therefore
$$\begin{aligned} TG &> GA \\ &> GH. \end{aligned}$$

It follows that the triangle FTG is greater than half the area $TEAH$.

Similarly, if the arc AH be bisected and the tangent at the point of bisection be drawn, it will cut off from the area GAH more than one-half.

Thus, by continuing the process, we shall ultimately arrive at a circumscribed polygon such that the spaces intercepted between it and the circle are together less than the excess of K over the area of the circle.

Thus the area of the polygon will be less than K.

Now, since the perpendicular from O on any side of the polygon is equal to the radius of the circle, while the perimeter of the polygon is greater than the circumference of the circle, it follows that the area of the polygon is greater than the triangle on any side of the polygon is equal to the radius of the circle, while the perimeter of the polygon is greater than the circumference of the circle, it follows that the area of the polygon is greater than the triangle K; which is impossible.

Therefore the area of the circle is not less than K.

Since then the area of the circle is neither greater nor less than K, it is equal to it.

Let us now put this proof into modern algebraic notation. We want to prove that

$$A = \tfrac{1}{2} rC.$$

Let $K = \dfrac{1}{2} rC$ (the area of the triangle). If $A \neq K$, then we have $A > K$ or $A < K$.

I. Suppose $A > K$. Inscribe a square, and let its side be s_1, its short radius be h_1, and its perimeter be p_1. The area of the square is

$$a_1 = \frac{1}{2} h_1 p_1 .$$

Now double the number of sides of the inscribed polygon, and keep on doubling it. For polygon n, the side is s_n, the short radius is h_n, and the perimiter is p_n; hence, the area is

$$a_n = \frac{1}{2} h_n p_n .$$

Now we have from the geometry of the situation that

$$h_1 < h_2 < \ldots < h_n < \ldots < r,$$

$$p_1 < p_2 < \ldots < p_n < \ldots < C,$$

and,

$$a_1 < a_2 < \ldots < a_n < \ldots < A.$$

Now choose N so that

$$A - a_N < A - \frac{1}{2} rC.$$

Then

$$\frac{1}{2} rC < a_N .$$

But since $h_N < r$, $p_N < C$, and $a_N = \frac{1}{2} h_N p_N$, we have

$$a_N < \frac{1}{2} rC,$$

a contradiction.

II. Suppose, on the contrary, $A < K$. Circumscribe a square, and let its perimeter be P_1; then the area is

$$A_1 = \frac{1}{2} r P_1 .$$

Now double the number of sides of the circumscribed figure, and keep doing it. If, for the nth polygon, the perimiter is P_n, then the area is

$$A_n = \frac{1}{2} r P_n .$$

From the geometry, we have

$$C < \ldots < P_n < \ldots < P_2 < P_1$$

and

$$A < \ldots < A_n < \ldots < A_2 < A_1 .$$

Choose N so that

$$A_N - A < \frac{1}{2} rC - A.$$

Then

$$A_N < \frac{1}{2} rC.$$

But $C < P_N$ and $A_N = \frac{1}{2} rP_N$, so

$$\frac{1}{2} rC < A_N ,$$

another contradiction.

It follows that $A = K = \frac{1}{2} rC$.

SUMMARY OF ARGUMENT. We are given that

$$h_1 < h_2 < \ldots < h_n < \ldots < r,$$

$$p_1 < p_2 < \ldots < p_n < \ldots < C < \ldots < P_n < \ldots P_2 < P_1 ,$$

and

$$\frac{1}{2} h_1 p_1 < \frac{1}{2} h_2 p_2 < \ldots < \frac{1}{2} h_n p_n < \ldots < A < \ldots < \frac{1}{2} rP_n < \ldots < \frac{1}{2} rP_2 < \frac{1}{2} rP_1 .$$

We want to prove

$$A = \tfrac{1}{2} rC.$$

The inequalities we are given imply that for every n,

$$\tfrac{1}{2} h_n \, p_n < \tfrac{1}{2} rC < \tfrac{1}{2} rP_n .$$

Furthermore, if $A \neq \tfrac{1}{2} rC$, then

(1) there is an N such that

$$A - \tfrac{1}{2} h_N \, p_N < \left| A - \tfrac{1}{2} rC \right| ;$$

(2) there is an M such that

$$\tfrac{1}{2} rP_M - A < \left| A - \tfrac{1}{2} rC \right| .$$

I. Suppose $A > \tfrac{1}{2} rC$. Then

$$\left| A - \tfrac{1}{2} rC \right| = A - \tfrac{1}{2} rC.$$

Hence, $\tfrac{1}{2} rC < \tfrac{1}{2} h_N \, p_N$, a contradiction.

II. Suppose $A < \tfrac{1}{2} rC$. Then

$$\left| A - \tfrac{1}{2} rC \right| = \tfrac{1}{2} rC - A.$$

Hence, $\tfrac{1}{2} rP_M < \tfrac{1}{2} rC$, another contradiction.

It follows that $A = \tfrac{1}{2} rC$.

SHORTER SUMMARY OF ARGUMENT. Give $\left| A - \tfrac{1}{2} rC \right|$ a name. Call it ε. Then we have:

(1) For each $\varepsilon > 0$, there is N such that

$$A - \tfrac{1}{2} h_N p_N < \varepsilon.$$

(2) For each $\varepsilon > 0$, there is N such that

$$\tfrac{1}{2} rP_N - A < \varepsilon.$$

Now let $\varepsilon > 0$ be given. Then there are N_1 and N_2 such that

(1) if $n \geq N_1$, $\quad A - \tfrac{1}{2} h_n p_n < \dfrac{\varepsilon}{2}$,

(2) if $n \geq N_2$, $\quad \tfrac{1}{2} rP_n - A < \dfrac{\varepsilon}{2}$.

Hence, if N is the maximum of N_1 and N_2, then

$$\tfrac{1}{2} rP_N - \tfrac{1}{2} h_N p_N < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This is true for any $\varepsilon > 0$. Now by the inequalities,

$$\tfrac{1}{2} h_N p_N < A < \tfrac{1}{2} rP_N,$$

and

$$\tfrac{1}{2} h_N p_N < \tfrac{1}{2} rC < \tfrac{1}{2} rP_N.$$

It follows almost immediately that $\left|A - \tfrac{1}{2}rC\right| < \varepsilon$ for every $\varepsilon > 0$, and hence $A = \tfrac{1}{2} rC$.

This is now a proof in modern limit theory! Perhaps students who have seen this will have an easier time understanding the basic definition.

Note that the last part of the argument has essentially the following form: given that for every n $A_n < a < B_n$ and $A_n < b < B_n$, and that for every $\varepsilon > 0$, there is an N such that for all $n \geq N$, $B_n - A_n < \varepsilon$, to prove $a = b$. To give a complete proof (using all the inequality rules in the right places) is a form of the original argument of Archimedes: suppose $a \neq b$. Then $a < b$ or $a > b$. If $a < b$, then let $\varepsilon = b - a$, and let $n \geq N$; we have $A_n < a < b < B_n$, and hence $\varepsilon = b - a < B_n - A_n$; this contradicts $B_n - A_n < \varepsilon$. The case for $a > b$ is symmetric. This means that modern limit theory includes in a sense the two-case proof by contradiction involved in the Greek method of exhaustion. On the other hand, modern limit theory puts this part of the argument in the same place each time, so that we do not constantly need to repeat it. The desire to avoid this repetition is not new: speaking of his own theory of limits (which was inadequate by our standards), Newton said, "These Lemmas are premised to avoid the tediousness of deducing involved demonstrations *ad absurdum*, according to the method of the ancient geometers." (See Struik (1969), p. 299.) Evidently, Newton and his contemporaries felt that proofs by the method of exhaustion were exhausting the mathematicians as well as the areas.

A similar transformation can be made with another example, Euclid XII, Proposition 2 (from Heath (1956)). Euclid begins the book with

PROPOSITION 1.

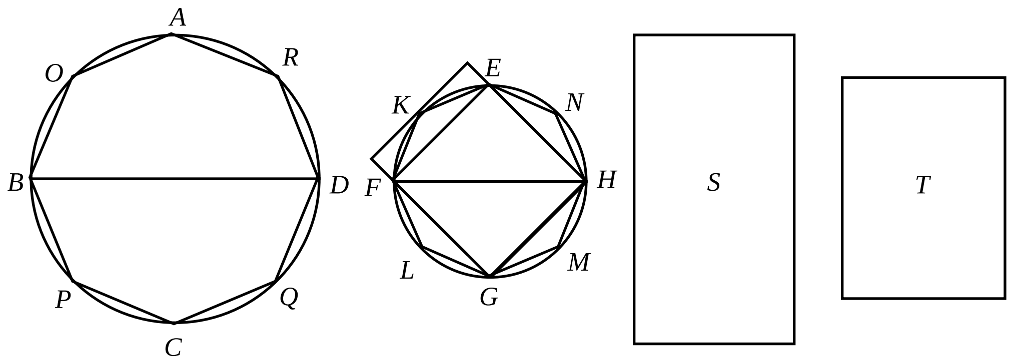
Similar polygons inscribed in circles are to one another as the squares on their diameters.

He then uses this to prove

PROPOSITION 2.

Circles are to one another as the squares on the diameters.

Let *ABCD*, *EFGH* be circles, and *BD*, *FH* their diameters;
I say that, as the circle *ABCD* is to the circle *EFGH*, so is the square on *BD* to the square on *FH*.



For, if the square on *BD* is not to the square on *FH* as the circle *ABCE* is to the circle *EFGH*,

then, as the square on *BD* is to the square on *FH*, so will the circle *ABCD* be either to some less area than the circle *EFGH*, or to a greater.

First, let it be in that ratio to a less area *S*.

Let the square *EFGH* be inscribed in the circle *EFGH*; then the inscribed square is greater than half of the circle *EFGH*, inasmuch as, if through the points *E*, *F*, *G*, *H* we draw tangents to the circle, the square *EFGH* is half the square circumscribed about the circle, and the circle is less than the circumscribed square;

hence the inscribed square *EFGH* is greater than the half of the circle *EFGH*.

Let the circumferences *EF*, *FG*, *GH*, *HE* be bisected at the points *K*, *L*, *M*, *N*,

and let *EK*, *KF*, *FL*, *LG*, *GM*, *MH*, *HN*, *NE* be joined;

therefore each of the triangles *EKF*, *FLG*, *GMH*, *HNE* is also greater than the half of the segment of the circle about it, inasmuch as, if through the points *K*, *L*, *M*, *N* we draw tangents to the circle and complete the parallelograms on the straight lines *EF*, *FG*, *GH*, *HE*, each of the triangles *EKF*, *FLG*, *GMH*, *HNE* will be half of the parallelogram about it,

while the segment about it is less than the parallelogram;

hence each of the triangles *EKF*, *FLG*, *GMH*, *HNE* is greater than the half of the segment of the circle about it.

Thus, by bisecting the remaining circumferences and joining straight lines, and by doing this continually, we shall leave some segments of the circle which will be less than the excess by which the circle *EFGH* exceeds the area *S*.

For it was proved in the first theorem of the tenth book that, if two unequal magnitudes be set out, and if from the greater there be subtracted a magnitude greater than the half, and from that which is left a greater than the half, and if this be done continually, there will be left some magnitude which will be less than the lesser magnitude set out.

Let segments be left such as described, and let the segments of the circle *EFGH* on *EK*, *KF*, *FL*, *LG*, *GM*, *MH*, *HN*, *NE* be less than the excess by which the circle *EFGH* exceeds the area *S*.

Therefore the remainder, the polygon *EKFLGMHN*, is greater than the area *S*.

Let there be inscribed, also, in the circle *ABCD* the polygon *AOBPCQDR* similar to the polygon *EKFLGMHN*;
therefore, as the square on *BD* is to the square on *FH*, so is the polygon *AOBPCQDR* to the polygon *EKFLGMHN*. [XII. 1]

But, as the square on *BD* is to the square on *FH*, so also is the circle *ABCD* to the area *S*;
therefore also, as the circle *ABCD* is to the area *S*, so is the polygon *AOBPCQDR* to the polygon *EKFLGMHN*; [V. 11]

But the circle *ABCD* is greater than the polygon inscribed in it;
therefore, the area *S* is also greater than the polygon *EKFLGMHN*.

But it is also less:
which is impossible.

Therefore, as the square on *BD* is to the square on *FH*, so is not the circle *ABCD* any area less than the circle *EFGH*.

Similarly we can prove that neither is the circle *EFGH* to any area less than the circle *ABCD* as the square on *FH* is to the square on *BD*.

I say next that neither is the circle *ABCD* to any area greater than the circle *EFGH* as the square on *BD* is to the square on *FH*.

For, if possible, let it be in that ratio to a greater area *S*.

Therefore, inversely, as the square on *FH* is to the square on *DB*, so is the area *S* to the circle *ABCD*.

But, as the area S is to the circle $ABCD$, so is the circle $EFGH$ to some area less than the circle $ABCD$;
therefore also, as the square on FH is to the square on BD, so is the circle $EFGH$ to some area less than the circle ABCD: [V. 11]
which was proved impossible.

Therefore, as the square on BD is to the square on FH, so is not the circle $ABCD$ to any area greater than the circle $EFGH$.

And it was proved that neither is it in that ratio to any area less than the circle $EFGH$;
therefore, as the square on BD is to the square on FH, so is the circle $ABCD$ to the circle $EFGH$.

Therefore etc.

Q. E. D.

LEMMA.

I say that, the area S being greater than the circle $EFGH$, as the area S is to the circle $ABCD$, so is the circle $EFGH$ to some area less than the circle $ABCD$.

For let it be contrived that, as the area S is to the circle $ABCD$, so is the circle $EFGH$ to the area T.

I say that the area T is less than the circle $ABCD$.

For since, as the area S is to the circle $ABCD$, so is the circle $EFGH$ to the area T,
therefore, alternately, as the area S is to the circle $EFGH$, so is the circle $ABCD$ to the area T. [V. 16]

But the area S is greater than the circle $EFGH$;
therefore the circle $ABCD$ is also greater than the area T.

Hence, as the area S is to the circle $ABCD$, so is the circle $EFGH$ to some area less than the circle $ABCD$. Q. E. D.

To put this argument into modern algebraic notation, let the given circles have areas a and b respectively, and let the corresponding ratio of the squares of their diameters be k. Let the areas of the polygons inscribed in the circle with area a have areas a_1 , a_2 , … Let the polygons inscribed in the other circle have areas b_1 , b_2 , … We have

$$0 < a_1 < a_2 < \dots a_n < \dots < a,$$

and

$$0 < b_1 < b_2 < \ldots b_n < \ldots < b.$$

Furthermore, for each n, we have

$$k = \frac{a_n}{b_n}.$$

In addition, we have for each n that

$$(a - a_{n+1}) < \tfrac{1}{2}(a - a_n), \qquad (b - b_{n+1}) < \tfrac{1}{2}(b - b_n).$$

We want to prove

$$k = \frac{a}{b}.$$

Now if $k \neq \frac{a}{b}$, then $k = \frac{a}{S}$, where $S < b$ or $S > b$.

I. Suppose $S < b$. Choose N so that

$$b - b_N < b - S.$$

Then

$$S < b_N.$$

But

$$S = \frac{a}{k} > \frac{a_N}{k} = b_N,$$

a contradiction.

II. Suppose $S > b$. This is similar to case I with a and b reversed.

It follows that

$$k = \frac{a}{b}.$$

REMARK. This proof depends on our being able to find an N so that $a - a_N < a - S$ or $b - b_N < b - S$. This, in turn, depends on

$$(a - a_{n+1}) < \tfrac{1}{2}(a - a_n), \qquad (b - b_{n+1}) < \tfrac{1}{2}(b - b_n).$$

From this we get easily

$$(a - a_{n+1}) < \left(\tfrac{1}{2}\right)^n (a - a_1), \qquad (b - b_{n+1}) < \left(\tfrac{1}{2}\right)^n (b - b_1).$$

In other words, $(a - a_n)$ and $(b - b_n)$ can be made as small as we please. Another way to say this is

(1) For each $\varepsilon > 0$, there is an N such that for all $n \geq N$,

$$|a - a_n| < \varepsilon.$$

(2) For each $\varepsilon > 0$, there is an N such that for all $n \geq N$,

$$|b - b_n| < \varepsilon.$$

SHORTER PROOF. Let $\varepsilon > 0$ be given. Choose N so that for all $n \geq N$,

$$|b - b_n| < \frac{1}{|k|}\varepsilon.$$

Then

$$|kb - a_n| = |kb - kb_n| = |k||b - b_n| < \varepsilon.$$

Hence, kb is the limit of the sequence a_n, and since a sequence can have only one limit, $kb = a$.

Note that this gives us a proof that the limit of kb_n is k times the limit of b_n.

REFERENCES

Heath, Thomas L. (1912) *The Works of Archimedes*. Reprinted by Dover (no date).

Heath, Thomas L. (1956) *The Thirteen Books of Euclid's Elements*, Dover, New York.

Landau, Edmund (1960) *Differential and Integral Calculus*, translated by Melvin Hausner and Martin Davis. Chelsea, New York. Originally published as *Einführung in die Differentialrechnung und Integralrechnung*, 1934.

Struik, D. J. (1969) *A Source Book in Mathematics, 1200–1800*, Harvard University Press, Cambridge, MA.