Differentials and Higher Order Differentials

from Leibniz to Euler

by

H. J. M. Bos

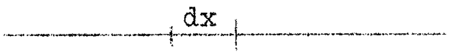
Text of a lecture of 1972; not published.

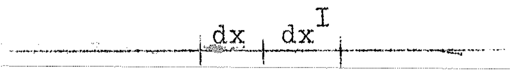
Differentials and Higher Order Differentials from Leibniz to Euler*

Ever since the seventeenth century the differential has haunted both mathematics and history of mathematics. In a collection of articles from the American Mathematical Monthly, published only three years ago, there is a section discussing how to introduce the differential in teaching. At the end of the section the conclusion is reached that

> "there is no commonly accepted definition of differential which fits all uses to which the notation is applied." (1)

And I am told (2) that a Canadian Ph.D. student has been able to distinguish no less than seven different types of infinitesimals in the history of mathematical analysis. I am sure that the questions and the arguments in the debate over differentials are quite familiar to you: Are differentials equal or unequal to zero? If unequal, are they infinitely small or finite but approaching zero? If infinitely small, what does that mean and will that notion not always lead to contradictions? If differentials are finite but approaching zero, do they ever become zero? And if not why can powers of the differentials be discarded? If the differentials are zero, how can you divide by them? Or do differentials simply not exist? Are they found only in symbols standing for mathematical entities arrived at by well defined processes? If that is so, why are they so suggestive? etc. I will not go into these questions because I have found them of little help in understanding the mathematics to which most of my historical studies have been directed. I have been studying the infinitesimal analysis practised on the continent by Leibniz and those mathematicians who, in the late seventeenth and eighteenth centuries, developed the differential and integral calculus in the way which Leibniz had initiated. It seems to me that only in rare occasions the course which their researches took can be explained by a concern for foundational questions concerning the differential. What I do find, however, is that there are certain practical and conceptual questions concerning the differential, and especially the higher order differential, which were of direct concern for that school of mathematicians. These questions have not received the attention they deserve, because historians of mathematics have usually restricted their attention to the foundational questions around the first order differential.

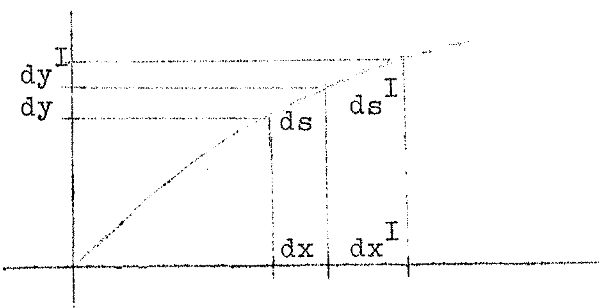




To introduce these questions, I invite you to look beyond the first order differential. What does one see when one looks beyond the first order differential? One sees the next first order differential. This differential is important because it occurs in the definition of the second order differential:

$$ddx \text{ (or } d^2x) = dx^I - dx$$

Now, what is dx^I? Especially, what is it with respect to dx? They are both differentials, so the obvious thing to suppose is that they are equal. In that case ddx would be zero; there would be no higher order differentials. Indeed, if we only look at the X-axis, the domain of the variable x, there is no reason to suppose otherwise and higher order differentials do not make sense. This shows that we must not restrict ourselves to the X-axis and study the differential as separate entity, but that we must look at the problem situation in which differentials occur.



The paradigm object of enquiry for the Leibnizian differential calculus is the curve, drawn with respect to perpendicular axès. This is important to stress: the calculus is an analytical calculus for geometrical problems. The famous book by the Marquis de l'Hôpital is most significantly called <u>Analyse des infiniment petits, pour l'intelligence des lignes courbes</u> (3) and I regret very much that usually the last part of the title of this first textbook of the calculus is omitted.

Now, let us come back to the differential and the next differential and ask: are they equal? Well, they may or they may not be; but surely, if

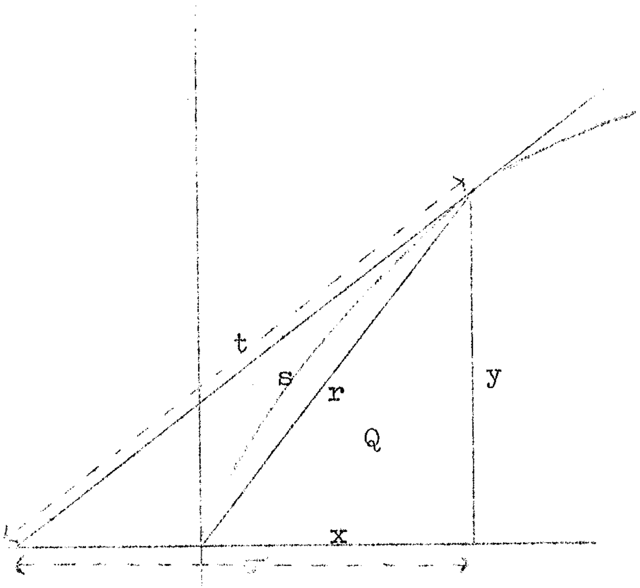
$$dx^{I} = dx$$

then $$dy^{I} \neq dy,$$

for otherwise the curve would have no curvature.

So we cannot suppose all first order differentials equal. But we may suppose the dx's equal. That choice would give a privileged position to the X-axis in the problem situation. In fact, as will appear later, the choice of supposing all dx's equal is in a certain sense equivalent to treating all the variables involved as functions of x; that is to take x as independent variable.

However, in the geometrical problems to which the Leibnizian infinitesimal analysis was applied, the X-axis did not have a privileged position. The curve was not considered as the graph of a function $x \rightarrow y(x)$, in which x is the independent variable. In fact, in the analysis of curves, until well into the eighteenth century, the concept of function as mapping, as unidirectional relation between an independent variable and a dependent variable, was absent.



Variables:

x abscissa
y ordinate
s arclength
r radius
t tangent
σ subtangent
Q quadrature

The curve was conceived as embodying a set of relations between variable geometrical quantities. Such variable geometrical quantities are for instance: abscissa, ordinate, arclength, radius, tangent, subtangent, quadrature, etc. The relations between these variables were expressed in the Analysis, if possible, by equations.

There is an essential difference between these variable geometrical quantities, or variables for short, and functions. The concept of variable does <u>not</u> imply the choice of one special variable in the problem to be considered as independent and as the variable on which all the other variables depend.

To illustrate this distinction between variables and functions we may turn to physical theories. Such theories are usually concerned with relations between variables ($pV = RT$ for instance), no particular one of which is an "independent" one. If we consider a process of fall, there are three variables; space traversed <u>s</u>, time <u>t</u> and velocity <u>v</u>. These are variables; there is no reason why velocity, for instance, should depend on time rather than on space traversed. In fact, if one introduces functions in this situation, the one variable velocity dissolves into two entirely different functions <u>v(t)</u> and <u>v(s)</u>.

The absence of the concept of function in the Leibnizian infinitesimal analysis and the predominance of the concept of variable is of crucial importance, because it explains why the derivative does not occur, and could not occur, in that calculus. Indeed, to define a derivative, the specification of the independent variable is necessary, and in the absence of a specified independent variable a derivative cannot be introduced. Variables cannot have derivatives, but they can have differentials. This fact explains also why it has been so difficult entirely to get rid of the differentials - as some rigorists in mathematics have deeply wanted to do. Differentials tend to be very persistent, especially in those cases where variables are involved, that is, in physics or applied mathematics generally.

But if functions, and hence derivatives, were absent in the Leibnizian calculus, the important question to be asked must be: when and why did derivatives eventually come to take over the role of fundamental concept in the differential calculus? The usual answer to this question is that the derivative emerged gradually in the eighteenth century, that it was canonised in the works of Lagrange and Cauchy and that the reason why the derivative took over, was that the logical inconsistencies of the differential became so embarrassing that one had to go over to the ratio, or the limit of the ratio, of two differentials. But I have personally always felt a little doubtful about this explanation. After all, would not the most typical reaction of the working mathematician to the foundational problem of the differential be to pay lip-service to it in the prefaces of his works, and to continue to work with differentials? This, in fact, was what Euler did and yet in his work the derivative - or to be precise the differential coefficient - plays a most important role.

Indeed, there were other reasons for the emergence of the derivative. The emergence of the function concept itself must be taken into account, and also the study of functions of more than one variable brought in the derivatives, because the usual conceptions and techniques of differentials break down when applied to these functions, and the ensuing technical problems naturally force one to the derivatives, in this case the partial derivatives.

As you see, these are technical and conceptual, rather than foundational problems, and in the rest of my talk I will concentrate on yet another reason for the emergence of the derivative, which also lies in the sphere of the conceptual and the practical. This reason is connected with higher order differentials.

As is obvious from my preceding arguments, the Leibnizian calculus, lacking the concept of derivative, had to introduce higher order differentials in order to deal with problems involving higher order differentiation. Yet, unlike the hardy first order differentials, the higher order differentials have been almost completely abolished from mathematics. It is reasonable to suppose, that the technical and conceptual difficulties associated with higher order differentials were so severe that they had to be eliminated. I shall argue that this is indeed the case, and that the attempt to eliminate higher order differentials was one of the main causes of the emergence of the derivative.

To introduce the aspects of higher order differentials which I have in mind, I ask you to return to the figures which we left for our discussions on functions and variables. We have seen that there is in principle no privileged variable, and therefore no a priori reason why we should suppose the dx's all equal, or, for instance, the dy's or the ds's. On the other hand, the dx's may be all equal - as long as we don't expect the other differentials to be all equal too. So there is an essential element of arbitrariness in the relation between adjacent differentials.

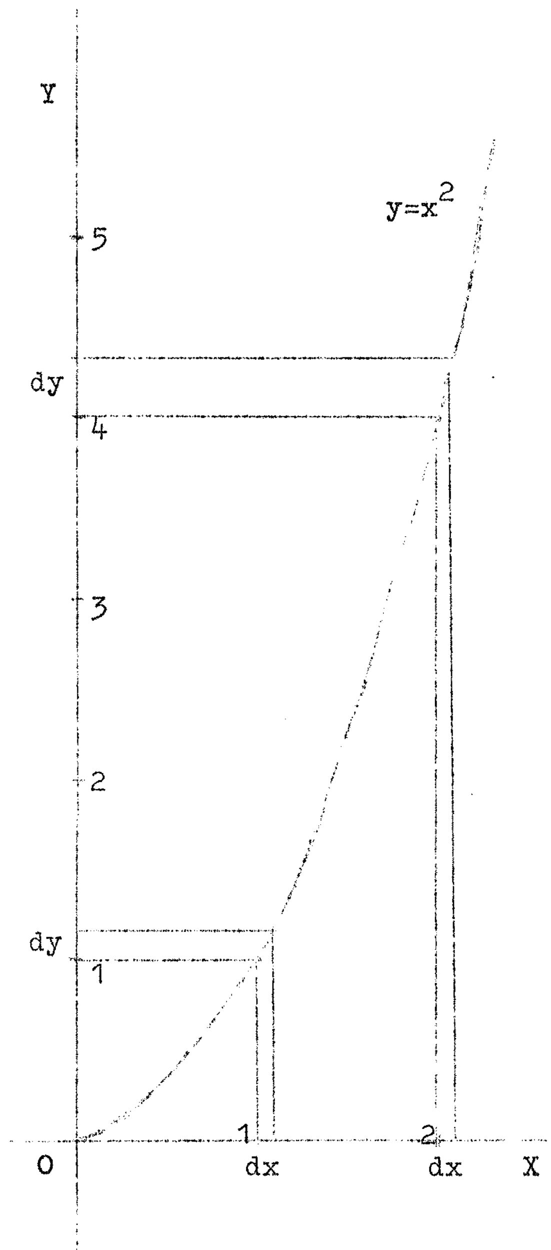
In order to discuss this point further I have to introduce a technical term. I shall refer to the way in which the successive differentials of the variables are related, as partition. If all the dx's are equal I shall say that the X-axis has a regular partition - and similarly for the Y-axis and the domains of the other variables. The practitioners of the Leibnizian calculus did not use this term; they used circumlocutions involving the idea of a progression of values of the variable (5).

We have seen that the partition has a degree of arbitrariness. Of course if the partition of the domain of one variable is fixed, the partitions of the other domains are fixed too. Thus the different possible partitions for a given curve can be specified by stating how the differentials of one variable behave. The four most usual partitions which occur in the works of Leibniz, the Bernoulli's and Euler are:

dx constant	regular partition of the X-axis
dy constant	regular partition of the Y-axis
ds constant	regular partition of the curve
ydx constant	regular partition of the area under the curve.

At this point I have to stop for two side remarks.

The partition tells us how the differential behaves along the domain of the relevant variable. If for instance for the parabola $y = x^2$ (see figure) dx is taken constant, the dy (for x = 2) is twice the dy (for x = 1). That is, the differential can be considered throughout the domain. Thus the differential becomes a variable, like the other variables in the problem situation, only infinitely small. It is, I feel, the failure of historical studies to see that the differential is an infinitely small variable, that has blocked much of our understanding of the differential. (The fact that one differential may be supposed constant is not at variance with its status as variable. Indeed, constant variables occur in many situations, as for instance: the constant ordinate of a horizontal straight line, the constant radius of curvature of a circle and the constant subtangent of the logarithmic curve).



My second remark concerns the question how these arguments about the partition relate to the formula which is at present still in use for the second derivative, d^2y/dx^2.

For $y = f(x)$, the derivative is, as you know, defined by $\frac{dy}{dx} = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$

The second derivative is usually introduced as the derivative of the derivative. However, one can also introduce it as

$$\frac{d^2y}{dx^2} = \lim_{h\to 0} \frac{[f(x+2h) - f(x+h)] - [f(x+h) - f(x)]}{h^2}$$

which is analogous to

$$\frac{d^2y}{dx^2} = \frac{dy^I - dy}{dx^2} \quad .$$

However, as you can see, for this definition of the second derivative it is essential that one takes the two segments h along the X-axis equal.

Indeed, let us try to introduce the second derivative directly as a limit of a quotient of finite differences with respect to segments h_1 and h_2 along the X-axis, which are not necessarily equal. The numerator of the quotient would be

$$[f(x+h_1+h_2) - f(x+h_1)] - [f(x+h_1) - f(x)].$$

But we run into a problem of choice for the denominator for which we might equally well chose h_1^2 or h_2^2 or, as a compromise, h_1h_2. But, for whatever choice of the denominator, the double limit for $h_1\to 0$, $h_2\to 0$ would not exist, as can be checked easily in the example $y = x$. So we have to suppose $h_1 = h_2$, which is equivalent to what in Leibnizian terminology is rendered as supposing dx constant. So we conclude : only if dx is taken constant does d^2y have a relation to the second derivative of $y(x)$. Put more generally, the variable whose differential is supposed constant takes a role equivalent of that of the independent variable.

Well, after all this, one should ask: does it matter? Obviously as long as we restrict ourselves to first order differentials it does not matter whether we suppose dx constant or dy constant etc. because these suppositions refer to adjacent differentials and they do not enter the story. So we state:

Problems involving only first order differentials do not depend on the partition.

But it does matter for higher order differentials. I shall illustrate this with some examples. The radius of curvature suggests itself as an example here because it involves repeated differentiation. Let me present you a series of formulas which Jakob Bernoulli gave in 1694(6) for the radius of curvature:

$$r = \frac{dxds}{ddy} \quad \text{(ds constant)} \qquad ; \qquad r = \frac{ds^3}{dxddy} \quad \text{(dx constant)}$$

$$r = \frac{ds^3}{dyddx} \quad \text{(dy constant).}$$

So we see:

The formula for a mathematical entity involving repeated differentiation depends on the partition chosen.

Conversely, if you have a formula involving higher order differentials its meaning depends on the partition. Let us return to Bernoulli's formula and apply it in the case of a circle $x^2 + y^2 = a^2$ but under different suppositions about the partition. We find

$$\frac{dxds}{ddy} = a \text{ for ds constant (as expected)}$$

$$= -y^2/a \quad \text{for dx constant}$$

$$= \infty \quad \text{for dy constant.}$$

So: The meaning of a formula involving higher order differentials depends on the partition.

To complicate matters still further consider the formula

$$\frac{dyds^2}{ddxds - ddsdx} .$$

Now if you work this formula out with respect to different partitions you find that you will always get the same result. In fact, this is a formula for the radius of curvature which is independent of the partition; Leibniz gave it, be it in a slightly different form, in 1694 (7), and he stated that, compared with Bernoulli's formulas, it has the advantage of being independent of the partition. So we conclude

Some formulas involving higher order differentials are indpendent of the partition.

Obviously, these aspects of higher order differentials are of importance for higher order differential equations. Consider e.g.

$$addx = (dy)^2 \quad ds \text{ constant} .$$

In the correspondence between Leibniz and Johann Bernoulli there is (dated 1694) a reference to this differential equation (8). However, the specification of the partition was omitted so that it reads there

$$addx = (dy)^2.$$

When I first was confronted with this differential equation I was greatly puzzled by the importance which the correspondents attached to it, because I thought it obvious that

$addx = (dy)^2$ yields $d^2x/dy^2 = 1/a$, whence

$x = y^2/2a + my + n$ would be the solution.

That is, I interpreted the differential equation as applying under the condition dy constant. However, properly interpreted,

$$addx = (dy)^2 \qquad ds \text{ constant}$$

has as solution $y = \pm a \arcsin(me^{-x/a}) + n$, a solution which was given by Jakob

Bernoulli in 1693 (be it not in this analytical form but by means of a geometrical construction of the solution curves) (9).

So here we see that the dependence of differential formulas on the partition implies that the solution of higher order differential equations depend on the partition with respect to which they are considered.

Reviewing these examples we see that the higher order differentials have a peculiar feature which the first order differentials lack: they involve an indeterminacy. The first order differential is affected by a logical inconsistency but routine manipulations with them leads to no ambiguities or pitfalls. But when we turn to higher order differentials, even in ordinary mathematical practice anomalies occur because a formula with higher order differentials can mean anything or nothing, depending on the assumed partition.

How much were the practitioners of the Leibnizian calculus aware of this indeterminacy? Very much. Leibniz comments on it several times, he stresses the necessesity to specify the partition when dealing with higher order differentials, and he claims - and rightly so - that the possibility to assume whatsoever partitions of the X-axis is precisely what makes his calculus superior to Cavalieri's (10). For lack of time I will not go into the studies of the Bernoulli's concerning the question of indeterminacy; suffice it to say that these aspects of higher order differentials belonged to the standard knowledge of the practitioners of the Leibnizian calculus.

Of these practitioners, Euler was the man who most fully grasped the implications of the indeterminacy of higher order differentials. He devoted large parts of his famous textbook <u>Institutiones calculi differentialis</u> (1755) (11) to the discussion of these implications. Not only did Euler see the implications of the indeterminacy of higher order differentials, he also concluded that these higher order differentials, because of their indeterminacy, do not really belong to Analysis. His argument was this (12): If the partition is not specified, higher order differentials are vague and have no determined meaning, so they do not belong to Analysis - unless, of course, they occur in formulas which are independent of the partition. But in that case, Euler says, they effectively cancel each other, so that they do not <u>really</u> enter Analysis. If, on the other hand, the partition <u>is</u> specified, the higher order differentials can be expressed in terms of finite variables and powers of first order differentials, so that in that case also they do not really enter Analysis. Indeed, if a partition is specified, this means that the differential of a certain variable is supposed constant, say dt constant. But then for every variable x one can form $dx=pdt$, in which p is a finite variable called the differential coefficient. Similarly $dp = qdt$ and $dq = rdt$. Now by means of these differential coefficients p, q, r, etc., the higher order differentials of x can be eliminated. Indeed $d^2x = d(pdt) = dpdt = qdt^2$, $d^3x = rdt^3$ etc. Thus the higher order differentials are eliminated by reducing them to powers of the first order differential dt and to the differential coefficients. The differential coefficients p, q, r etc. as well as the differential dt are independent of the partition, because their definitions involve only first order differentials.

What is important in this procedure is that by introducing the differential coefficients Euler arrives at formulas which are independent of the partition. In other practitioners of the calculus, found that in order to achieve independence of the partition one has to introduce differential coefficients.

These problems about the indeterminacy of higher order differentials are not foundational problems of the same sort as those about first order differentials. They are practical - specification of the partition is part of the mathematical argument - and they are conceptual - for instance where they concern symbols whose meaning is not fixed. We see that these problems find a natural solution in the introduction of differential coefficients with respect to one variable. That is, they induce a consideration of the variables involved as functions of one specified variable, in which case the differential coefficients become the derivatives. This then concludes my argument that the indeterminacy of higher order differentials was one important contributing factor to the emergence of the derivative.

The emergence of the derivative accompanied the disappearance of higher order differentials and arbitrary partitions. To end this talk I would like to give an example of how these aspects may still pop up in modern mathematical practice.

The formula for the second derivative, d^2y/dx^2, is a formula which depends on the partition. If we consider for example the function $y = x^2$, then

$$d^2y/dx^2 = 2 \text{ for } dx \text{ constant},$$

$$d^2y/dx^2 = 0 \text{ for } dy \text{ constant},$$

$$d^2y/dx^2 = 2(1 - 4x^2)^{-1} \text{ for } ds \text{ constant},$$

$$d^2y/dx^2 = -2 \text{ for } ydx \text{ constant.}$$

Thus d^2y/dx^2 is one of those formulas to which the practitioners of the Leibnizian calculus would always add an indication of the partition. We don't. We have even forgotten that we should. This is not too serious because we usually interpret d^2y/dx^2 as a single symbol and we know that y should be considered as a function of x.

But problems may arise, for instance when we consider the chain rule for first derivatives:

$$\frac{dy}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt} ,$$

in which, apparently, the dx's in the right hand side cancel each other. Applying the same procedure to second derivatives we arrive at

$$\frac{d^2y}{dt^2} = \frac{d^2y}{dx^2}\left(\frac{dx}{dt}\right)^2 , \qquad (*)$$

which formula is incorrect; it should be

$$\frac{d^2y}{dt^2} = \frac{d^2y}{dx^2}\left(\frac{dx}{dt}\right)^2 + \frac{dy}{dx} \cdot \frac{d^2x}{dt^2} . \qquad (**)$$

So we may ask: why is it that apparently we are allowed to cancel dx's in the case of first order derivatives but not in the case of higher order derivatives?

The answer is that in fact we may cancel dx's as in (*), but that we are in general not allowed to interpret d^2y/dx^2 and d^2y/dt^2 as second derivatives of y(x) and y(t) respectively. In the Leibnizian calculus, (*) is a perfectly correct equality between differentials, but the meaning of the terms on both sides of the equation depends on the partition. If dx is chosen constant then d^2y/dx^2 may be interpreted as the second derivative of y(x), but d^2y/dt^2 may not be interpreted as the second derivative of y(t); it is a quotient of differentials, for whose evaluation one has to express numerator and denominator in terms of dx, that is, one has to consider both y and t as the functions of x. Conversely, if dt is supposed constant, d^2y/dt^2 may be interpreted as the second derivative of y(t), but d^2y/dx^2 may not be interpreted as the second derivative of y(x). Hence the only case that (*) may be interpreted as an equality between derivatives is when dx and dt can be taken constant together, that is, when x is linearly dependent on t. But this also follows from the chain rule for derivatives (**), because if x is linearly dependent on t, the second term of the right hand side of (**) vanishes, so that (**) reduces to (*).

NOTES

*This is a slightly altered and anotated version of a paper read to the British Society for the History of Mathematics during its conference at Nottingham on 20th May, 1972.

(1) Selected papers on the calculus, ed. T.M. APOSTOL e.a., Mathematical Association of America, 1969, p. 186.

(2) By dr I. Grattan-Guinness in private correspondence.

(3) Paris, 1696.

(4) In his Institutiones Calculi Differentialis, St. Petersburg, 1755, reprinted in EULER, L. Opera Omnia, Ser. I, Vol. X, Leipzig-Berlin, 1913.

(5) So for instance Leibniz:

> Es ist ganz nicht nöthig ad summandum, dass die dx oder dy constantes und die ddx = 0 seyen, sondern man assumiret die progression der x oder y (welche man pro abscissa halten will) wie man es gut findet. (LEIBNIZ, G. W. Mathematische Schriften, ed. C. I. GERHARDT, Berlin-Halle, 1849-1863, reprint Hildesheim 1961-1962, Vol. VII, p. 387)

and Euler:

> . . . notavimus differentias secundas atque sequentes constitui non posse, nisi valores successivi ipsius x certa quadam lege progredi assumantur; quae lex cum sit arbitraria. . .
>
> (op. cit. note (4) part I paragraph 128)

(6) Acta Eruditorum, June 1694, reprinted in Jak. BERNOULLI, Opera, Geneva, 1744, p. 576-600.

(7) Acta Eruditorum August 1694, reprinted in Mathematische Schriften (see note (5)) Vol. V, p. 309-318. In fact, Leibniz gave $r = \frac{dx}{d(dy/ds)}$; working out $d(dy/ds) = (ddyds - ddsdy)/(ds)^2$, one arrives at $r = dyds^2/(ddyds-ddsdy)$. Euler presented several formulas which involve higher order differentials but are independent of the partition, in his Institutiones (see note (4)); such as

$$\frac{dyddx - dxddy}{dx^3} \quad \text{(Part I, paragraph 257)},$$

and

$$\frac{(dx^2 + dy^2 + dz^2)^{3/2}}{(dx+dz)ddy - (dy+dz)ddx + (dx-dy)ddz} \quad \text{(Part I, paragraph 261)}.$$

(8) Johann Bernoulli to Leibniz 19th May 1694, see LEIBNIZ Mathematische Schriften Vol. III p. 139.

(9) Acta Eruditorum, June 1693, reprinted in Jakob BERNOULLI, Opera p. 549-573.

(10) Compare for instance:

> . . . aream figurae calculo meo designo $\int ydx$, seu summam ex rectangulis cujusque y ducti in respondens sibi dx, ubi si dx ponantur se aequales, habetur Methodus indivisibilium Cavalerii. (LEIBNIZ, mss Elementa Calculi novi, published in GERHARDT C.I. Die Geschichte der höheren Analysis, erste Abteilung, die Entdeckung der höheren Analysis, Halle 1855, pp 149-155, esp. p. 150)

> Und das ist eben auch der avantagen meines calculi differentialis, dass mann nicht sagt die summa aller y, wie sonst geschehen, sodern die summa aller ydx oder $\int ydx$, denn so kan ich das dx expliciren und die gegebene quadratur in andere infinitis modis transformiren und also eine vermittelst der andern finden. (Mathematische Schriften, Vol. VII p. 387)

Sed haec (that is Cavalieri's) Indivisibilium Methodus tantum initia quaedam ipsius artis continebat (. . .). Nam quoties ordinatim ductae inter se parallelae, nempe rectae lineae vel planae superficies (. . .) intercipiunt inaequalia quaedam elementa, non licet ipsas ordinatim applicatas in unum addere, ut contentum figurae prodeat, sed ipsa intercepta Elementa infinite parva sunt mensuranda; (. . .). Ea vero infinite parvorum aestimatio Cavalerianae methodi vires excedebat . . . (LEIBNIZ, mss Scientiarum diversos gradus . . ., published by G. I. GERHARDT in Monatsberichte der Könichlich Preussischen Akademie der Wissenschaften zu Berlin (28 October) 1875, pp 595-599, esp. p. 597.)

(11) See note (4). The most important chapter in this respect is Ch. VIII of part I, entitled De formularum differentialium ulteriori differentiatione

(12) Part I, paragraph 263 sqq.