What really is a continuous function?

J.F. Harper,
School of Mathematics, Statistics and Computer Science,
Victoria University, Wellington, New Zealand.
e-mail john.harper@vuw.ac.nz
phone (+64)(4)463 5662
fax (+64)(4)463 5045

1   Introduction

When teaching elementary analysis a few years ago from Bartle and Sherbert (1992), I was amazed to discover that the definition of continuity for a function from the real line $\boldsymbol{R}$ to $\boldsymbol{R}$ was different from the one I had learnt as a student, using Hardy's "Pure Mathematics" (10th ed., 1952). Further investigation showed that even expert authors had trouble with continuity: two authors (Hardy and Whittaker) changed their definitions after writing their first editions, Whittaker and Watson (1927) is unclear, and Goursat and Hobson contain an error that was known to be an error at the time.

Ten textbooks published over the last century or so contain at least five (possibly six) different definitions of a continuous function of a real variable. The intention of this paper is to prove this surprising fact, by quoting the authors' definitions and showing what they imply for the continuity at $x = 0$ of three functions f, g, h defined to be zero at every point in their respective domains $F, G, H \subset \boldsymbol{R}$, and to be undefined elsewhere:

$$F = \{0\}, \text{ the set whose only member is 0;}$$
$$G = \boldsymbol{Q}, \text{ so } g \text{ is undefined if } x \text{ is irrational;}$$
$$H = [0, \infty), \text{ so } h \text{ is undefined if } x < 0.$$

The ten books' definitions of continuity apply to those functions as follows:

Jordan (1893)	f, g, h
Harkness and Morley (1893 [1925 reprint])	none
Whittaker (1902)	none
Goursat (1904)	h
Pierpont (1905)	g, h
Hobson (1907)	g
Hardy (1908)	none
Whittaker and Watson (1927)	h or f, h ?
Hardy (1952)	h
Bartle and Sherbert (1992)	f, g, h.

It had been suggested to me that something was odd about the teaching of analysis in late nineteenth-century and early twentieth-century Cambridge, where many British mathematicians learnt their subject. The problem about continuity is not, however, purely Cambridge-generated: the six authors who were educated there (Harkness, Morley, Whittaker, Hobson, Hardy, Watson) have three, or possibly four, essentially different definitions between them. The five authors educated elsewhere were Goursat and Jordan (French), Pierpont (American, with a Vienna PhD), and Bartle and Sherbert (both American), and they have three different definitions, one of which also appears among the Cambridge group.

One error occurs in two of our authors, one from Cambridge (Hobson) and one not (Goursat). Hedrick, whose English translation of Goursat was published in 1904, gave a footnote correcting the error. Either Hobson (1907) had read Goursat in the original French, or he perpetrated the same error independently, or they both got it from the same source which I failed to find.

Hardy said in his preface that he was writing for bright first-year undergraduates intending to specialise in mathematics; Bartle and Sherbert said they were writing for mathematicians in their first Real Analysis course; Goursat's translator said that that book could be used conveniently in the American system of instruction as a text for a *second course in calculus* (italics here and elsewhere are in the original); all the other books in the list seem to have been for more advanced students.

2 The Definitions

The ten books have at least five different opinions about what constitutes a continuous function of a real variable. That surprising conclusion arises from finding which of our three functions f, g, h are continuous at $x = 0$ according to the definitions given. (Checking all this would be a good exercise for students of either analysis or history of mathematics!)

Jordan (1893): f, g, h all continuous at $x = 0$

p46 "Soit $f(x, y, \dots)$ une fonction des n variables $x, y, \dots$ définie dans une ensemble E.

"Soient $(a, b, \dots)$ un point déterminé de E; $h, k, \dots$ des quantités variables, assujetties à la seule condition que le point $(a+h, b+k, \dots)$ appartienne aussi à E.

"Si, pour tout valeur de la quantité positive ϵ, on peut determiner une autre quantité positive δ, telle que l'on ait

$$|f(a + h, b + k, \dots) - f(a, b, \dots)| < \epsilon$$

pour tous les systèmes de valeurs de $h, k, \dots$ pour lesquels on a

$$|h| < \delta, \qquad |k| < \delta, \qquad \dots,$$

on dira que la fonction $f(x, y, \dots)$ est *continue au point* $(a, b, \dots)$."

My translation of that is as follows.

"Let $f(x, y, \dots)$ be a function of the n variables $x, y, \dots$ defined in a set E.

"Let $(a, b, \dots)$ be a fixed point of E; $h, k, \dots$ variable quantities, subject to the one condition that the point $(a + h, b + k, \dots)$ belongs also to E.

"If, for any value of the positive quantity ϵ, we can find another positive quantity δ, such that we have

$$|f(a + h, b + k, \dots) - f(a, b, \dots)| < \epsilon$$

for all the systems of values of $h, k, \dots$ for which we have

$$|h| < \delta, \qquad |k| < \delta, \qquad \dots,$$

we shall say that the function $f(x, y, \dots)$ is *continuous at the point* $(a, b, \dots)$."

Harkness and Morley (1893): none of f, g, h continuous at $x = 0$.

p42 "[Throughout this chapter and the next, ϵ will be used, always, to denote an arbitrarily small positive number]."

p49 "The function $f(x)$ is said to be continuous at the point c ... if a field $(c - h$ to $c + h)$ can be found such that for all points of this field, $|f(x) - f(c)| < \epsilon$."

Note: I cannot find any definition of "field" on or before p49, but I assume it means "interval" in this context.

Hardy (1908): none of f, g, h continuous at $x = 0$.

p172 "We must first define continuity *for any particular value of x*. Let us fix on some particular value of x, say the value $x = \xi$... What are the characteristic properties of $\phi(x)$ associated with this value of x?

"In the first place $\phi(x)$ *is defined for* $x = \xi$....

"Secondly $\phi(x)$ *is defined for all values of x near* $x = \xi$; i.e. we can find an interval, including $x = \xi$ in its interior, for all points of which $\phi(x)$ is defined.

"Thirdly *if x approaches the value ξ from either side $\phi(x)$ approaches the limit $\phi(\xi)$*...

"DEFINITION. *The function $\phi(x)$ is said to be continuous for $x = \xi$ if it tends to a limit as x tends to ξ from either side, and each of those limits is equal to $\phi(\xi)$*.

"...our definition is equivalent to '$\phi(x)$ *is continuous for $x = \xi$ if, given ϵ, we can choose η so that $|\phi(x) - \phi(\xi)| < \epsilon$ if $0 \leq |x - \xi| \leq \eta$*).'"

He emphasises his definition in p175 no.22: "The function which is equal to 1 when x is rational and to 0 when x is irrational (Ch. II, Ex. XVII. 11) is discontinuous for all values of x. So too is any function which is defined only for rational or for irrational values of x."

Hardy (1952): h continuous at $x = 0$, f, g not.

He still includes everything quoted above from his 1908 edition (except that η, ϵ have become ϵ, δ respectively, and the reference to Ch. II, Ex. XVII. 11 has been altered as needed), but he also says

p186 "We have often to consider functions defined only in an interval (a, b). ... We shall say that $\phi(x)$ is continuous for $x = a$ if $\phi(a+0)$ exists and is equal to $\phi(a)$." (He had already defined $\phi(a + 0)$ as the one-sided limit, and his notation (a, b) means a closed interval.)

4

Goursat (1904): h continuous at $x = 0$, f, g not.

Note: "the interval (a, b)" appears to mean the closed interval for Goursat. On p2, after supposing that x "can assume all values between two given numbers a and b $(a < b)$." he says "Let y be another variable, such that to each value of x between a and b, and also for the values a and b themselves, there corresponds one definitely determined value of y. Then y is called a function of x, defined in the interval (a, b); and this dependence is indicated by writing the equation $y = f(x)$....

p2 "Let $y = f(x)$ be a function defined in a certain interval (a, b), and let x_0 and $x_0 + h$ be two values of x in that interval. If the difference $f(x_0 + h) - f(x_0)$ approaches zero as the absolute value of h approaches zero, the function $f(x)$ is said to be *continuous for the value x_0*.

p6 "The function $y = x \sin 1/x$, for example, is a perfectly continuous function of x, for $x = 0$,* and y approaches zero as x approaches zero." The very necessary footnote was added by the translator, E.R. Hedrick, Professor of Mathematics in the University of Missouri:

"* After the value zero has been assigned to y for $x = 0$.—TRANSLATOR."

(The translator's preface to the English edition says inter alia "To [Professor Goursat] is due all the additional matter not to be found in the French text, except the footnotes which are signed, and even these, though not of his initiative, were always edited by him.")

Whittaker (1902): none of f, g, h continuous at $x = 0$.

p41 "Let $f(z)$ be a quantity which, for all values of z lying within given limits, depends on z.

"Let z_1 be a point situated within those limits. Then $f(z)$ is said to be *continuous* at the point z_1 if, corresponding to any positive quantity ϵ, however small, a finite positive quantity η can be found, such that the inequality

$$|f(z) - f(z_1)| < \epsilon$$

is satisfied so long as $|z - z_1|$ is less than η." [The discussion preceding this quote shows that Whittaker had in mind both $\boldsymbol{R}$ and $\boldsymbol{C}$.]

Whittaker & Watson (1927): none of f, g, h is continuous at $x = 0$ by their p42 definition, but h is by their p44 extension, and so is f if $a = b$ is permitted in their definition of "simple curve", which would allow a simple curve to consist of one point. It is not clear to me whether they did allow that.)

p42 "Let $f(x)$ be a function of x defined when $a \leq x \leq b$. Let x_1 be such that $a \leq x_1 \leq b$. If there exists a number l such that, corresponding to the arbitrary positive number epsilon, we can find a positive number η such that

$$|f(x) - l| < \eta,$$

whenever $|x - x_1| < \eta$, $x \neq x_1$, and $a \leq x \leq b$, then l is called the limit of $f(x)$ as $x \to x_1$.

"It may happen that we can find a number l_+ (even when l does not exist) such that $|f(x) - l_+| < \epsilon$ when $x_1 < x < x_1 + \eta$. We call l_+ the limit of $f(x)$ when x approaches x_1 from the right and denote it by $f(x_1 + 0)$; in a similar manner we define $f(x_1 - 0)$ if it exists.

"If $f(x_1 + 0)$, $f(x_1)$ and $f(x_1 - 0)$ all exist and are equal, we say that $f(x)$ is *continuous* at x_1."

p44 "Let $f(z)$ be a function of z defined at all points of a closed region (one- or two-dimensional) in the Argand diagram, and let z_1 be a point of the region. Then $f(z)$ is said to be continuous at z_1 if, given any positive number ϵ, we can find a corresponding positive number η such that $|f(z) - f(z_1)| < \epsilon$, whenever $|z - z_1| < \eta$ and z is a point of the region." Their "closed one-dimensional region" is a simple curve, defined on p43 (using the p42 definition of continuity) by "Let x and y be two functions of a real variable t which are continuous for every value of t such that $a \leq t \leq b$. We denote the dependence of x and y on t by writing

$$x = x(t), \qquad y = y(t). \qquad (a \leq t \leq b)$$

The functions x and y are supposed to be such that they do not assume the same pair of values for any two different values of t in the range $a < t < b$. Then the set of points with coordinates (x, y) corresponding to these values of t is called a simple curve."

Pierpont (1905): g, h continuous at $x = 0$; "limiting point" on p208 excludes f.

Note: in this book, $\Re$ means the set of all real numbers, and $\Re_m$ means what one would normally think of as $\Re^m$.)

p208 "Let $f(x_1 \ldots x_m)$ be defined over a domain D. Let $a = (a_1 \ldots a_m)$ be a proper limiting point of D. If

$$\lim_{x=a} f(x_1 \ldots x_m) = f(a1 \ldots a_m),$$

the function f is continuous at a ... The reader should observe that a is not only a limiting point of D, but that it lies in D." The definition of "proper limiting point" is on p157,158.

"Let A be a point aggregate in $\mathfrak{R}_m$. Any point p of $\mathfrak{R}_m$ is a limiting point of A, if however small $\rho > 0$ is taken, $D_\rho(p)$ contains an infinity of points of A. If every domain of p contains at least one other point, p is a limiting point of A ... If p is a limiting point of A and p itself lies in A, it is called a proper limiting point."

And on p153 $D_\rho(p)$ was defined to be the points in R_m whose distance from p is less than or equal to ρ. Pierpont's definition of limit needs quoting as it's not the usual ϵ–δ one:

p171 "Let $f(x)$ be a one-valued function defined over a domain D. Let

$$A = a_1, a_2, a_3, \ldots$$

be *any* sequence of points in D such that

$$\lim a_n = a; \qquad a \text{ finite or infinite}; \qquad a_n \neq a.$$

If the sequence $f(a_1), f(a_2), f(a_3), \ldots$ has a limit η, finite or infinite, always the same, however the sequence A be chosen, we say ... η is the limit of $f(x)$ for $x = a$ ..."

Hobson (1907): g continuous at $x = 0$, h excluded by p222 quotation, f excluded by p222-3 quotation.

p221 "Let the domain of the independent variable x be continuous, and either bounded or unbounded; and denote the function y at the point x by $f(x)$.

"The function $f(x)$ is said to be continuous at the point α of the domain of x, if, corresponding to any arbitrarily chosen positive number ϵ whatever, a positive number δ dependent on ϵ can be found, such that $|f(\alpha+\eta)-f(\alpha)| < \epsilon$, for all positive or negative values of η which are numerically less than δ, and which are such that $\alpha + \eta$ is in the domain of x. At an an end-point of a limited domain, the values of η will have one sign only."

p222 "A function which is not continuous at a point α may satisfy the condition that in a neighbourhood of α on the right the fluctuation of the function may be made as small as we please ... the function is then said to be *continuous on the right* at α."

p222-3 "The domain of the independent variable has hitherto been considered continuous; it is however clear from a consideration of the definition of continuity ...that the definition is applicable in case the domain of the independent variable is not continuous, but consists of any set of points which contains limiting points that belong to the set. It is, of course, only at such a limiting point that the question of continuity arises ...the notion of continuity of a function is applicable whetever be the domain ...except when it consists of an isolated set of points."

p61 "If a linear set of points not finite in number (denoted by G) is in the interval (a, b), then a point P, in whose arbitrarily small neighbourhood there exists at least one point of G not identical with P, is called a limiting point of the set G, whether P belongs to G or not."

Note that Hobson p236 has a wrong example, which is the same as Goursat's if $a = 0$: "Let $f(x) = (x-a) \sin 1/(x-a)$; then $f(a+0) = 0, f(a-0) = 0$. This function is continuous at x = a. . . " He must have imagined $k(x)l(x)$ to be defined at 0 if $k \to 0$ as $x \to 0$ and l is bounded in a deleted neighbourhood of $x = 0$.

Bartle and Sherbert (1992): (f, g, h all continuous at $x = 0$.)

p140 "Let $A \subseteq \mathbf{R}$, let $f : A \to \mathbf{R}$, and let c be in A. We say that f is continuous at c if, given any neighbourhood $V_\epsilon(f(c))$ of $f(c)$ there exists a neighbourhood $V_\delta(c)$ of c such that if x is any point of $A \cap V_\delta(c)$, then $f(x)$ belongs to $V_\epsilon(f(c))$."

Neighbourhoods were defined on p41: $V_\epsilon(a) = \{x \in \mathbf{R} : |x - a| < \epsilon\}$.

3 Conclusions

Eleven experts on analysis defined continuity in ways that give five or six different results, two of the authors committed an error, and one famous book does not resolve an ambiguity. Perhaps we should not be too hard on our students if they fail to appreciate the subtleties.

4 References

Bartle, R.G. and Sherbert, D.R. 1992 *Introduction to real analysis*, 2nd ed. Wiley, New York.

Goursat, E. 1904 *A course in mathematical analysis*, English trans. by E.R. Hedrick. Ginn, Boston.

Harkness, J. and Morley, F. 1893, *A treatise on the theory of functions.* Stechert, New York.

Hardy, G.H. 1908 *A Course of Pure Mathematics* 1st ed. Cambridge University Press.

Hardy, G.H. 1952 *A Course of Pure Mathematics* 10th ed. Cambridge University Press.

Hobson, E.W. 1907 *The Theory of Functions of a Real Variable and the Theory of Fourier's Series.* Cambridge University Press.

Jordan, M.C. 1893 *Cours d'Analyse de l'Ecole Polytechnique*, Tome 1, 2nd ed. Gauthier-Villars, Paris.

Pierpont, J. 1905 *Lectures on the Theory of Functions of Real Variables*, Vol. 1. Ginn, Boston.

Whittaker, E.T. 1902 *A Course of Modern Analysis.* Cambridge University Press.

Whittaker, E.T. and Watson, G.N. 1927 *A Course of Modern Analysis* 4th ed. Cambridge University Press.