Some critical notes on the Cantor Diagonal Argument

Philip Molyneux

This paper critically examines the Cantor Diagonal Argument (CDA) that is used in set theory to draw a distinction between the cardinality of the natural numbers and that of the real numbers. In the absence of a verified English translation of the original 1891 Cantor paper from which it is said to be derived, the CDA is discussed here using a consensus from the forms found in a range of published sources (from "popular" to "professional"). Some general comments are made on these sources. The discussion then focusses on the CDA as applied to the correspondence between the set of the natural numbers, and the set of real numbers in the open range (0,1) in their expansion from decimal digits (0.123… etc.).

Four points critical of the CDA are raised: (1) The conventional presentation of the CDA forms a putative new real number \(X\) from the "diagonal" of the chosen list of real numbers and which is therefore not on this initial list; however, it omits to consider that it may indeed be on the later part of the list, which is never exhausted however far the "diagonal" list is extended. (2) This aspect, combined with the fact that \(X\) is still composed of decimal digits, that is, it is a real number as defined, indicates that it must be on the later part of the list, that is, it is not a "new" number at all. (3) The conventional application of the CDA leads to one putative "new" real number \(X\); however, the logical extension of this in its "exhaustive" application, that is, by using all possible different methods of alteration of the decimal digits on the "diagonal", and by reordering the list in all possible ways, leads to a list of putative "new" real numbers that become orders of magnitude longer than the chosen "diagonal" list. (4) The CDA is apparently considered to be a method that is applicable generally; however, testing this applicability with the natural numbers themselves leads to a contradiction.

Following on from this, it is found that it indeed is possible to set up a one-to-one correspondence between the natural numbers and the real numbers in (0,1), that is, \(\mathbb{N} \leftrightarrow \mathbb{R}\); this takes the form: … \(a_j a_k a_i \leftrightarrow 0.a_j a_k a_i\), where the right hand extension of the natural number is intended to be a mirror image of the left hand extension of the real number. It is also shown how this may be extended to real numbers outside the range (0,1).

Additionally, a form of the CDA was presented by Wilfred Hodges in his 1998 critical review of "hopeless papers" dealing with the CDA; this form is also examined from the same viewpoints, and to the same conclusions.

Finally, some comments are made on the concept of "infinity", pointing out that to consider this as an entity is a category error, since it simply represents an absence, that is, the absence of a termination to a process.

1. Introduction

1.1. The concept of infinity is evidently of fundamental importance in number theory, but it is one that at the same time has many contentious and paradoxical aspects. The current position depends heavily on the theory of infinite sets and the concept of one-to-one correspondence that was introduced over a century ago by the German mathematician Georg Cantor [2-4]. In essence, the Argument aims to show that a "new" real number can be produced which differ from those in a list of the real numbers when indexed against the natural numbers, so that this former list is not exhaustive.
1.2. Fundamentally, any discussion of this topic ought to start from a consideration of the work of Cantor himself, and in particular his 1891 paper [3] that is presumably to be considered the starting point for the CDA.

1.3. In fact, with Cantor's 1891 paper [3], the relevant text - at page 76 in the reference - shows that he considered here specifically an infinite set with two types of elements (m and w) in a specific order. This is unlike the present format of the CDA as conventionally presented and as discussed below, since there no numbers involved.

1.4. However, it seems we lack a validated and accepted English translation of this paper [3], with clarification of the technical terms used at that time by the author and by contemporary mathematicians, although such translations are available for other Cantor papers [4,10,16]. Since the present publication is in English, we are forced to fall back on the English literature to deal with the subject.

1.5. Although the format of the method used here may not follow that of Cantor, it follows that of the published treatments, which range from what may be called "popular" accounts to those from demonstrably "professional" sources [5-9, 11, 13, 15, 17-19]. In fact, any distinction between "popular" and "professional" presentations is not sharp, and not strictly relevant to the present discussion. These published accounts use methods that differ in detail, and in the details of the procedure applied. The present discussion uses the conventional approach to the formulation of the CDA applied to real numbers, as their decimal expansion in the open interval (0,1).

1.6. It is useful, as points of comparison for the later discussion, to examine the way in which infinite processes are normally treated, considering two elementary examples: the one-to-one mapping between the whole set of the natural numbers and the set of the even natural numbers; and the summation of a converging infinite series.

1.7. In the case of one-to-one mapping between the natural numbers \( \mathbb{N} \) and the even natural numbers \( \mathbb{N}' \), we have a sequence of the form

\[
\begin{align*}
\mathbb{N} & \quad \mathbb{N}' \\
1 & \leftrightarrow 2 \\
2 & \leftrightarrow 4 \\
3 & \leftrightarrow 6 \\
4 & \leftrightarrow 8 \\
etc. & \quad etc.
\end{align*}
\]

At each stage we have a positive outcome, in the sense that the pairing produced supports the contention of a one-to-one mapping of the two series, and there is nothing to let us believe that this conclusion would alter if the process were extended indefinitely. Indeed, this may be treated without any consideration of infinity, since in the above listing each of the entries has the form: \( n' = 2n \), so that the one-to-one correspondence follows naturally.

1.8. As the second example, consider the process the summation of a converging infinite series, for example:

\[
S = 1/2 + 1/4 + 1/8 + \ldots
\]

where the limit of the series is of course unity; here, as the summation is continued, the difference of the sum from unity exactly may be made as small as required by having enough terms, and this may be satisfied however small this difference is specified to be (the "epsilon-delta criterion"). Here again, a doubting questioner may be satisfied in a suitable way eventually, while the validity can be demonstrated in by the standard mathematical proof (\( 2S = 1 + 1/2 + \ldots = 1 + S \), etc.). In any case, there is
no reason to doubt that there would be any sudden change in this conclusion as the process is continued indefinitely.

2. Basis of the CDA

2.1. To consider the CDA, the sources already cited [5-9, 11, 13, 15, 17-19] show diverse formats. In particular, some formats use symbols, and other use actual digits, in the expansion of the real numbers. It is simplest to consider one particular format, and then look at the differences of other formats that have been used. In fact, the format used here does not follow exactly that of the published examples, but is typical in most senses.

2.2. The format to be used is then that of the natural numbers $\mathbb{N}$ paired in succession with the rational numbers $\mathbb{R}$ in the open interval $(0,1)$, it being assumed that this process does eventually exhaust all the natural numbers in some way.

2.3. The only requirement is that the rational numbers in this interval may be presented by infinite decimals, each of the digits in which is between 0 and 9 (that is, in this case we are using base 10).

2.4. A typical example for the first three stages of this would be as below, where certain numerals in the expansions of $\mathbb{R}$ are in bold type for the purposes of the CDA:

<table>
<thead>
<tr>
<th>$\mathbb{N}$</th>
<th>$\mathbb{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.123...</td>
</tr>
<tr>
<td>2</td>
<td>0.456...</td>
</tr>
<tr>
<td>3</td>
<td>0.789...</td>
</tr>
<tr>
<td>etc.</td>
<td>etc.</td>
</tr>
</tbody>
</table>

2.5. The CDA then proceeds by constructing a putatively new number by taking the diagonal formed by the digits in bold and then altering them in some specified way, such as by adding one in a cyclic fashion (0→1, ..., 9→0). This then produces a number designated here as $X$:

$$? \quad 0.260... = X$$

where the indexing number shown as $?$ is evidently different from all those listed. This process may evidently be continued indefinitely, and since the new number is different from all of those listed, then it is seemingly it is a "new" number that has no matching natural number.

2.6. The conventional conclusion that is then drawn is that there are more numbers on the list of real numbers $\mathbb{R}$ than there are on the list of natural numbers $\mathbb{N}$.

2.7. In the alternative case of the use of symbols, the sequence is interrupted at the general natural number $n$

<table>
<thead>
<tr>
<th>$\mathbb{N}$</th>
<th>$\mathbb{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0.a_{11} a_{12} a_{13} \ldots a_{1n} \ldots$</td>
</tr>
<tr>
<td>2</td>
<td>$0.a_{21} a_{22} a_{23} \ldots a_{2n} \ldots$</td>
</tr>
<tr>
<td>3</td>
<td>$0.a_{31} a_{32} a_{33} \ldots a_{3n} \ldots$</td>
</tr>
<tr>
<td>etc1</td>
<td>etc1</td>
</tr>
<tr>
<td>$n$</td>
<td>$0.a_{n1} a_{n2} a_{n3} \ldots a_{nn} \ldots$</td>
</tr>
<tr>
<td>etc2</td>
<td>etc2</td>
</tr>
</tbody>
</table>
(where the distinction between the "etc1." and "etc2." is discussed below in ¶4.2) and the new number is formulated as

\[ \begin{align*}
? & \quad 0. b_1 \ b_2 \ b_3 \ldots b_n \ldots = B
\end{align*} \]

where the diagonal argument is ensured by making

\[ b_n \neq a_{nn} \]

with the same conclusion being made as in ¶2.6.

3. Preliminary comments on the CDA formats

3.1. It is useful to make some preliminary comments on the CDA formats used in the published literature.

3.2. Considering the definition used in the formulation the real numbers in the interval (0.1), as already noted in any ¶2.3, any listing with the denary system (base ten) implicitly defines the sequence of entries after the decimal point as:

First entry: Digit between 0 and 9
Second entry: Digit between 0 and 9
Third entry: Digit between 0 and 9
etc. etc.

3.3. Considering the format used for the real numbers, in most cases of the methods published where digits rather than symbols are used, the real numbers are presented in the form: 0.12345…; however, this is itself an infinity of numbers, since the termination “…" may have 0 to 9 as the first digit, 0 to 9 as the second digit, and so on. In the present case this might be avoided by putting the numbers as ending in an indefinite sequence of zeroes: 0.123450. Here the underlining is used on the last digit as a notation to indicate an indefinite sequence of that same digit that is typographically easier than the conventional form of a dot above the number. However, it is simpler to keep to the consensus in the published sources, and to accept any ambiguity in this termination “…”.

3.4. The choice of the sequence of the real numbers in the range 0 to 1 in the published methods is often apparently arbitrary, and such a listing is therefore not evidently exhaustive. This applies whether the listing is given as actual digits, or as symbols. Following on the previous comments, it might therefore be preferable to list the real numbers (0,1) as mirroring the respective real number, which would ensure the real number list is complete:

\[ \begin{align*}
\mathbb{N} & \quad \mathbb{R} \\
1: & \quad 0.10000 \\
2: & \quad 0.2\underline{0000} \\
3: & \quad 0.30000 \\
4: & \quad 0.4\underline{0000} \\
\text{etc.} & \quad \text{etc.}
\end{align*} \]

where numbers highlighted in bold form the diagonal as discussed below. In the later entries the format would be, for example

\[ \begin{align*}
10: & \quad 0.01\underline{0000000000}
\end{align*} \]

with added decimal places as necessary. However, since this apparent lack of completeness is not thought a defect even in the "professional" presentations [7, 8, 15, 16], it will also be passed over here.
3.5. Where digits (rather than symbols) are used in the published examples, the numbers of places of decimals quoted vary widely - from three by Enderton [8], five by Hofstadter [13], eight by Gamow [9], ten by Penrose [18], to eleven in the anonymous author in the current Wikipedia [1] (who also uses binary rather than denary notation). It may be noted that Enderton in his professional text [8] takes the method to be so obvious that he does not restrict himself to the interval (0,1) but instead chooses three rather wildly chosen numbers: 236.001…, -7.777…, and 3.1415….

3.6. These differences in the number of listed numbers also appear where symbols are used [5, 7, 15] although here this may be rendered irrelevant by the use of the general entry for natural number \( n \) as in §2.7 above.

3.7. These differences in the choice of the number of entries in the natural/real listing may show some uncertainty in the CDA itself, that is, more entries may make the method seem stronger.

4. Two critiques of the CDA

4.1. Two specific critiques are given here, which are independent in their approach, but concur in their results. Two other critiques from other viewpoints are also given in §§5 and 6 below.

4.2. In the first critique, we examine the conclusion that is drawn from the production of the putative new natural number. It is stated that since this is different from any number on the list of natural numbers, then this list is not exhaustive. However, there is a caveat that is omitted here, which is that this only applies to numbers in the finite list as written. For there is an infinite number of entries below those specified, as indicated rather casually by the "etc." in §2.4 for examples where digits are used, and by the "etc2." in §2.7 where symbols are used. These represent real numbers still to be examined, to select a different entry at the specific place in the decimal expansion. This means that however this far this process may continued, at each stage there always remains an infinite number not yet considered.

4.3. Thus the only proper conclusion to be drawn this stage, is that it cannot be excluded that the "diagonal number" \( X \) is on the remainder of the list, and that that applies however far the specified section of the list is extended.

4.4. By contrast, it seems that in the conventional approach, the construction of the "diagonal number" is viewed as a completed process that has been performed instantaneously - although however short the (finite) time to perform each step, it still requires an infinite number of steps to finish the process. However, the conclusions as given here in §4.3 - that the "diagonal number" could be on the remainder of the list - continue to apply at each stage in this process, but this seems to be abrogated at the assumed end of the process.

4.5. This abrogation is in contrast to parallel "infinite" processes in mathematics, as in the correspondence between the set of all the natural numbers and the set of the even natural numbers (§1.4), or the summation of a convergent series (§1.5), where in each case there is no abrogation of the conclusions drawn were the processes to be carried to completion.

4.6. The second critique involves considering the "diagonal number" \( X \) in the light of the definition of the real numbers given above in §2.2 and again in §3.2. For its format shows that it fulfills the criterion that each entry in the number is a digit between 0 and 9. This in turn shows that this fulfills the definition of a real number between 0 and 1, and hence one that must be on the list, albeit lower down.

4.7. These critiques support one another, in that the uncertainty of the conclusion from the first critique - which indeed allows the "diagonal number" \( X \) to be further down
the list - is resolved by the second critique - which shows that it must be present further down the list. Taken together, these critiques therefore cast doubt on the strength of the Cantor Diagonal Argument.

5. Exhaustive application of the CDA

5.1. In its conventional formulation, the CDA is applied only once to obtain one "diagonal number" \( X \). However, this restriction is arbitrary, and it is necessary to examine the outcome of an exhaustive application of the CDA to the present example, allowing any change in the digits on the "diagonal" (§5.2), and allowing any order for entries in the list (§5.3).

5.2. The conventional approach leads to only one putatively "new" number, replacing one digit in each entry by a different one as specified. However, for completeness, consider all the cases where each digit in the "diagonal" is changed to one of the nine other possible digits. For the "diagonal" list of three entries in §2.4, there are \( 10 \times 10 \times 10 = 1000 \) different ways of choosing the 10 digits in the "diagonal", with 1 of these being the original one, so that the exhaustive application of the CDA leads to 999 "new" numbers; in the general case, with \( n \) entries, this would be \( n^{10} - 1 \). Whatever the case, a list of "new" numbers is produced which is greater than that in the original list of real numbers according to a power factor. However, this depends on the number system used, so that with the binary system there is indeed only one way of producing the "new" number. Nevertheless, the choice of the number base here is arbitrary, and by choosing a larger base the number of possibilities may be increased accordingly.

5.3. Additionally, and for further completeness, the entries in the "diagonal" list may presumably be ordered in any way, each giving another number \( X \) from the diagonal process. Thus, with the listing of just three numbers in §2.4 above, this gives \( 3! = 6 \) ways of ordering the numbers, and 6 ways of producing the number \( X \) even from a single replacement method with the binary system \((0 \rightarrow 1, \text{etc.})\). In the general case, with \( n \) entries, there would be \( n! \) ways of ordering the list, and hence \( n! \) of the numbers \( X \) rather than just one.

5.4. Thus when the CDA is applied not just once, but exhaustively in these ways, this would give a list of valid derived numbers \( X \) that becomes extensively longer than the chosen "diagonal" list itself.

6. Application of the CDA to the natural numbers

6.1. Since the CDA is presumably a general argument, it is useful to see the result if it is applied to the natural numbers themselves.

6.2. Again this is presented in two columns, with column \( \mathbb{N} \) being the indexing column, and column \( \mathbb{N}' \) being the natural numbers in the format as extended in this case to the right with sufficient zeroes to allow the CDM to be applied:

\[
\begin{array}{c|c}
\mathbb{N} & \mathbb{N}' \\
1 & 00001 \\
2 & 00002 \\
3 & 00003 \\
4 & 00004 \\
\text{etc.} & \text{etc.}
\end{array}
\]
In this case the underlining on the right indicates the zeroes extend indefinitely to the right. The bold digits are again those to be used in the CDA.
6.3. If the CDA is applied as before, then this leads as before to a putatively “new” natural number Y:

\[ ? \quad 01112 = Y \]

However, it is clear that this is again a natural number, to be found further down the list.
6.4. This test of the CDA on the natural numbers themselves leads to a contradiction, which again suggests that is some doubt in using the CDA in more general applications [13].

7. A one-to-one correlation between the natural numbers and real numbers

7.1. The earlier discussion has cast some doubt on the conclusion currently drawn for the CDA, that it is not possible to produce a one-to-one correspondence between the natural numbers and the real numbers. However, from some clues provided by this discussion, particularly from the format of the natural numbers \( \mathbb{N} \) that has been used in \( \S 6.2 \), it turns out that it is possible to produce such a correspondence.
7.2. Consider, for example, the following sequence referring to the decimal expansion for \( \pi/10 \):

\[
\begin{array}{c|c}
\mathbb{N} & \mathbb{R} \\
\hline
3 & 0.3 \\
13 & 0.31 \\
413 & 0.314 \\
1413 & 0.3141 \\
etc. & etc.
\end{array}
\]

leading therefore eventually to:

\[ …951413 \leftrightarrow 0.314159… \]

where the “…” on the right hand side is to be read as the remainder of the decimal expansion of \( \pi/10 \), and that on the left hand side is to be read as the mirror image of this expansion.
7.3. It is evident that the entries in \( \mathbb{R} \) column may be made as close as required to the decimal expansion of \( \pi/10 \), while the entries in column \( \mathbb{N} \) still remain natural numbers. The limit of this process therefore provides a one-to-one correspondence between the two forms. Here the \( \mathbb{N} \)-numbers may be considering as derived from the \( \mathbb{R} \)-numbers either by reflecting them across the decimal point, or by rotating them by 180° about it.
7.4. This may be extended for real numbers not limited to the range (0,1). For example, for \( \pi \) itself, since it has a one-to-one correspondence with \( \pi/10 \), then this gives its one-to-one correspondence with the limit of this sequence. Here, to indicate the number of decimal places in the real number, we may place the corresponding number of zeroes after the natural number, so that the correlation for \( \pi \) itself becomes

\[
\begin{array}{c|c}
\mathbb{N} & \mathbb{R} \\
\hline
. & \pi
\end{array}
\]
with the correlate still being a natural number albeit in an unusual format. More generally, for example with $10^{17}\pi$, the correlated natural number could be denoted by \ldots 951413.0^{18}; here again the presence of the symbols after the decimal point does not detract from this being a natural number.

7.5. Using symbols rather than digits, this one-to-one correlation can be represented for the real numbers in the range $(0,1)$ as:

$$
\begin{array}{c}
\mathbb{N} \\
\mathbb{R}
\end{array} \quad \frac{\ldots a_n a_{n-1} a_1}{0.a_1 a_2 a_3 \ldots}
$$

where the same mirroring for the two forms of "\ldots" is taken to apply.

7.5. This leads us to the image of a game between two participants (mathematicians or otherwise) in which the first successively adds chosen digits to the left of the decimal expansion of the real number while the second adds the same digits to the right of the corresponding natural number, to the exhaustion of both participants.

8. Wilfred Hodges strictures on "hopeless papers"

9.1. Above any discussion of the CDA must loom the criticisms published by Wilfred Hodges in his 1998 review [12], arising from his experiences with "hopeless papers" during his work as an editor and as a referee/reviewer; these these criticisms were aimed specifically at the authors of submitted manuscripts, dealing with their objections to the CDA.

9.2. None of Hodges' quoted papers seem to deal with the critiques presented earlier here.

9.3. In this connection, Hodges presented his version of the CDA that he apparently considers to be authoritative (albeit, as Hodges says, "not in Cantor's own words"), as follows:

[Quotation starts]

(1) We claim first that for every map $f$ from the set $\{1, 2, \ldots\}$ of positive integers to the open unit interval $(0, 1)$ of the real numbers, there is some real number which is in $(0,1)$ but not in the image of $f$.

(2) Assume that $f$ is a map from the set of positive integers to $(0, 1)$.

(3) Write

$$
0. a_{n_1} a_{n_2} a_{n_3} \ldots
$$

for the decimal expansion of $f(n)$, where each $a_{n_i}$ is a numeral between 0 and 9. (Where it applies, we choose the expansion which is eventually 0, not that which is eventually 9.)

(4) For each positive integer $n$, let $b_n$ be 5 if $a_{n_n} \neq 5$, and 4 otherwise.

(5) Let $b$ be the real number whose decimal expansion is

$$
0 . b_1 b_2 b_3 \ldots
$$

(6) Then $b$ is in $(0,1)$.

(7) If $n$ is any positive integer, then $b_n \neq a_{n_n}$, and so $b \neq f(n)$. Thus $b$ is not in the image of $f$.

(8) This proves the claim in (1).

[Quotation ends]
(9) We deduce that there is no surjective map from the set of positive integers to the set \((0, 1)\).

(10) Since one can write down a bijection between \((0, 1)\) and the set of real numbers (and a bijection between the positive integers and the natural numbers, if we want the latter to include 0), it follows that there is no surjective map from the set of natural numbers to the set of real numbers.

(11) So there is no bijection between these two sets; in other words, they have different cardinalities.

[Quotation ends]

9.4. The present two critiques of §4 still apply to the proof in this format. For the second half of Hodges' text in his paragraph (7) in §9.3 only applies to the entries in the map \(f\) down to \(n\), with the later entries not having been examined to produce the next decimal places in the "new" number \(b\), so that this does not rule out this number being in this later part of the list (as noted already §4.2 above). Furthermore, the specification in Hodges (4) indicates that the number \(b\) will comprise a specific sequence of 4's and 5's, which will therefore be a real number (as noted already in §4.5 above).

10. The term “infinity”

10.1. The present paper has avoided as far as possible any consideration of the concept of infinity, although it is implicit in the sets of both the natural numbers and real numbers. However, it is useful to make some remarks here on this term.

10.2. The use of the term “infinity” as a noun, and even the use of a symbol for it (whether \(\infty\), \(\aleph_0\), or whatever), seems to imply that it is a defined entity or quantity, and one to which a comparative or even possibly a superlative could possibly be applied. However, this is evidently a category error, for the use of the alternative ("Anglo-Saxon") form of the word as “endlessness” reveals that it refers to the absence of a feature to a process, that is, to the absence of any limitation to the length of this process.

10.3. The parallels here are such terms as “blackness” (the absence of light in the environment, or the absence of reflectivity for a surface) or “vacuum” (the absence of material content) where again comparatives or superlatives evidently cannot be applied - one material object cannot be "blacker" than another, and one void cannot be more vacuous than another.

10.4. Such considerations in viewing "infinity" otherwise than is what the science-popularising author Lancelot Hogben may have meant when referred to this area as a "sematic quagmire" [14].

11. Concluding remarks

The previous discussion has focused on the CDA as the prime source for considering the cardinalities of the natural numbers and the real numbers, and has necessarily not considered other publications by Cantor where this distinction is dealt with. However, in so far as the sources cited represent the consensus of opinion on the validity of the CDA, the present criticisms suggest that the basis for consensus this needs to be reexamined.

REFERENCES
[3] G. Cantor, Jahresbericht der Deutschen Mathematiker-Vereinigung, 1891, 1, 75-78. [The title page of this (the first) volume of this journal is headed "1890-91" below the volume number, but then the book is dated "1892" at the bottom of the page, leading to some differences in this dating in the literature.]

Macrophile Associates, 33 Shaftsbury Avenue, Radcliffe-on-Trent, Nottingham, NG12 2NH, U.K.
E-mail address: drpmolyneux@zen.co.uk
03 August 2017