Georg Cantor and Transcendental Numbers

Robert Gray

1. INTRODUCTION. Conflicting statements have been made about Cantor's proof of the existence of transcendental numbers. For example, consider the following statements:

> The contrast between the methods of Liouville and Cantor is striking, and these methods provide excellent illustrations of two vastly different approaches toward proving the *existence* of mathematical objects. Liouville's is purely *constructive*; Cantor's is purely *existential*.
>
> —Mark Kac and Stanislaw M. Ulam [**20**, p. 13]

> Sometimes people have gone on to say that Cantor's method is not constructive, and cannot yield an explicit transcendental number. The words of E. T. Bell on page 569 of his *Men of Mathematics* are typical:
>
> > The most remarkable thing about Cantor's proof is that it provides no means whereby a single one of the transcendentals can be constructed.
>
> This is not true. Cantor's idea can be used so as to yield an utterly explicit transcendental number.
>
> —I. N. Herstein and I. Kaplansky [**19**, p. 238]

Kac and Ulam are comparing Cantor's method with Liouville's earlier construction of transcendentals [**24, 25**]—they find Cantor's method to be non-constructive. But Herstein and Kaplansky insist that Cantor's method is constructive, that it can produce a transcendental. A few lines later, they assert that this transcendental "is as well determined a number as e or π."

After reading the statements by Kac and Ulam, and Herstein and Kaplansky, we decided to study Cantor's work and how it has been presented. This article contains the results of our study. We begin by analyzing Cantor's original articles, his 1874 article that contains his first proof and his 1891 article that contains his diagonal proof. Our analysis will show that Cantor's methods lead to computer programs that generate transcendentals, and it will also determine which transcendentals are generated by the diagonal method. Next we will examine the history behind Cantor's first proof. Finally, we will consider how some commonly-held views about mathematics and its history have affected the interpretation of Cantor's work.

2. CANTOR'S FIRST PROOF. In 1874, Cantor published his first proof of the existence of transcendentals in an article titled "On a Property of the Collection of All Real Algebraic Numbers" [**3, 5**]. Cantor begins his article by defining the algebraic reals and introducing the notation: (ω) for the collection of all algebraic reals, and (ν) for the collection of all natural numbers. Next he states the property mentioned in the article's title; namely, that the collection (ω) can be placed into a one-to-one correspondence with the collection (ν), or equivalently:

> ...the collection (ω) can be thought of in the form of an infinite sequence:
>
> (2.) $$\omega_1, \omega_2, \ldots, \omega_\nu, \ldots$$
>
> which is ordered by a law and in which all individuals of (ω) appear, each of them being located at a fixed place in (2.) that is given by the accompanying index.

Cantor states that this property of the algebraic reals will be proved in Section 1 of his article, and then he outlines the rest of the article:

> To give an application of this property of the collection of all real algebraic numbers, I supplement Section 1 with Section 2, in which I show that when given an arbitrary sequence of real numbers of the form (2.), one can determine, in any given interval ($\alpha \cdots \beta$), numbers η that are not contained in (2.). Combining the contents of both sections thus gives a new proof of the theorem first demonstrated by Liouville: In every given interval ($\alpha \cdots \beta$), there are infinitely many transcendentals, that is, numbers that are not algebraic reals. Furthermore, the theorem in Section 2 presents itself as the reason why collections of real numbers forming a so-called continuum (such as, all the real numbers which are ≥ 0 and ≤ 1), cannot correspond one-to-one with the collection (ν); thus I have found the clear difference between a so-called continuum and a collection like the totality of all real algebraic numbers.

To appreciate the structure of Cantor's article, we number his theorems and corollaries:

Theorem 1. *The collection of all algebraic reals can be written as an infinite sequence.*

Theorem 2. *Given any sequence of real numbers and any interval* $[\alpha, \beta]$, *one can determine a number* η *in* $[\alpha, \beta]$ *that does not belong to the sequence. Hence, one can determine infinitely many such numbers* η *in* $[\alpha, \beta]$. (We have used the modern notation $[\alpha, \beta]$ rather than Cantor's notation ($\alpha \cdots \beta$).)

Corollary 1. *In any given interval* $[\alpha, \beta]$, *there are infinitely many transcendental reals.*

Corollary 2. *The real numbers cannot be written as an infinite sequence. That is, they cannot be put into a one-to-one correspondence with the natural numbers.*

Observe the flow of reasoning: Cantor's second theorem holds for *any* sequence of reals. By applying his theorem to the sequence of algebraic reals, Cantor obtains transcendentals. By applying it to any sequence that allegedly enumerates the reals, he obtains a contradiction—so no such enumerating sequence can exist. Kac and Ulam reason differently [**20**, p. 12–13]. They prove Theorem 1 and then Corollary 2. By combining these results, they obtain a non-constructive proof of the existence of transcendentals.

Cantor's theorems are worded constructively, but are they proved constructively? Since Cantor's original proof of Theorem 2 is not commonly known, we present it before answering this question.

Proof of Theorem 2: Recall that we have been given an interval $[\alpha, \beta]$ and a sequence of real numbers ω_n. We must find an η in $[\alpha, \beta]$ that does not belong to this sequence. Cantor assumes that the members of the sequence are distinct; to handle an arbitrary sequence, we can either eliminate duplicates from the sequence or modify his proof to handle arbitrary sequences.

Cantor begins his proof by finding the first two numbers in the given sequence that belong to $[\alpha, \beta]$. Denote the smaller of these numbers by α_1 and the larger by β_1. Now form the interval $[\alpha_1, \beta_1]$, and locate the first two numbers in the sequence that belong to $[\alpha_1, \beta_1]$. Denote the smaller of these numbers by α_2 and the larger by β_2. Then form the interval $[\alpha_2, \beta_2]$, and continue this procedure of generating intervals.

We have two cases: Cantor's procedure yields finitely many intervals $[\alpha_n, \beta_n]$ or infinitely many such intervals. In the first case, let $[\alpha_N, \beta_N]$ be the last interval generated. Since there can be at most one ω_k in $[\alpha_N, \beta_N]$, any η in the interval besides this ω_k and the endpoints of the interval will satisfy the conclusion of the theorem. In the second case, let $\alpha_\infty = \lim_{n \to \infty} \alpha_n$ and $\beta_\infty = \lim_{n \to \infty} \beta_n$. These limits exist since the α_n's form an increasing sequence that is bounded from above, and the β_n's form a decreasing sequence that is bounded from below.

The second case breaks into two cases: Either $\alpha_\infty = \beta_\infty$ or $\alpha_\infty < \beta_\infty$. At this point in his proof, Cantor notes that $\alpha_\infty = \beta_\infty$ holds for the sequence of algebraic reals [**3**, p. 261; **5**, p. 308–309]. So Cantor not only applies his theorem to the sequence of algebraic reals, but he also analyzes how his proof handles this particular sequence.

To complete the proof, we must produce a suitable η for the two remaining cases. In the case where $\alpha_\infty = \beta_\infty$, let η be this common limit. Note that η cannot be a member of the given sequence since for every k, ω_k does not belong to $[\alpha_{k+1}, \beta_{k+1}]$. In the case where $\alpha_\infty < \beta_\infty$, let η be any number in the interval $[\alpha_\infty, \beta_\infty]$.

Cantor's proof is constructive—he uses the given sequence ω_n to define the sequences α_n and β_n, breaks his argument into three cases depending on the behavior of these sequences, and constructs a suitable η for each case.* If the sequence ω_n contains all the algebraic reals, then α_n and β_n are converging nested sequences whose common limit η is a transcendental number.

Perhaps the most convincing way to show that Cantor's argument produces a transcendental is by computing one. Using the methods in Cantor's proof, we have written a computer program that generates the digits of a transcendental in the interval $(0, 1)$. Output from our program is given in Figure 1. Our program generates the sequence ω_n by enumerating the polynomials with integer coefficients and approximating their roots. We approximate roots by using Sturm's theorem and Horner's method [**31**, p. 138–156]. (For a precise description of the ω_n sequence generated by our program, see the appendix below.)

*We call a proof "constructive" if it constructs an object using methods acceptable to most mathematicians. For a proof of Cantor's Theorem 2 that meets the demands of constructive mathematicians, see [**2**, p. 27].

Approximation			Associated Polynomial
$\alpha_1 =$	ω_1	$= .50$	$2x - 1$
$\beta_1 =$	ω_2	$= .61\ldots$	$x^2 + x - 1$
$\alpha_2 =$	ω_{12}	$= .561\ldots$	$x^2 + 3x - 2$
$\beta_2 =$	ω_{19}	$= .577\ldots$	$3x^2 - 1$
$\alpha_3 =$	ω_{41}	$= .569\ldots$	$x^3 - x^2 + 2x - 1$
$\beta_3 =$	ω_{66}	$= .574\ldots$	$x^3 + 2x^2 + 2x - 2$
$\alpha_4 =$	ω_{87}	$= .57318\ldots$	$2x^3 - 2x^2 - 3x + 2$
$\beta_4 =$	ω_{359}	$= .57347\ldots$	$2x^3 + x^2 + 4x - 3$
$\alpha_5 =$	ω_{5539}	$= .573402\ldots$	$4x^4 - 3x^3 + 3x^2 + 2x - 2$
$\beta_5 =$	ω_{2159}	$= .573416\ldots$	$2x^4 - 4x^3 - 4x^2 - 2x + 3$
$\alpha_6 =$	ω_{156510}	$= .5734104\ldots$	$3x^5 + 5x^4 - x^3 + 4x^2 + 2x - 3$
$\beta_6 =$	ω_{144803}	$= .5734122\ldots$	$3x^5 + 3x^4 - 2x^3 - 3x^2 - 2x + 2$
$\alpha_7 =$	$\omega_{1406370}$	$= .57341146\ldots$	$x^6 - x^5 + 2x^4 + 3x^3 - x^2 + x - 1$
$\beta_7 =$	$\omega_{1057887}$	$= .57341183\ldots$	$x^6 - 4x^5 - x^4 + 5x^3 + 2x^2 + 3x - 3$

Figure 1. Generating a transcendental using Cantor's 1874 method.

While generating the ω_n sequence, our program also generates the sequences α_n and β_n, which approximate our transcendental η. Figure 1 shows the first seven members of α_n and β_n. As Cantor points out in his proof, these sequences have a common limit—so our program will produce closer approximations. In fact, we can calculate a bound on the number of algebraic reals that need examining in order to find an α_n and β_n that approximate η to within $1/k$. This calculation requires a look at our enumeration of the algebraic reals.

For ease of programming, we use a different enumeration of the algebraic reals than Cantor's. Cantor enumerates the polynomials with integer coefficients by their *height*, where the height of the polynomial $a_0x^k + \cdots + a_k$ is $k - 1 + |a_0| + \cdots + |a_k|$, and then he enumerates the roots of these polynomials. We enumerate polynomials by their *size*, which for the polynomial just mentioned is $\max(k, |a_0|, \ldots, |a_k|)$. Polynomials of the same size are ordered by treating them as $(k + 1)$-digit numbers whose digits range from $-k$ to k. For example, the polynomials $20x - 1$ and $x^{20} - 1$ are both size 20 and the first precedes the second in this ordering. Our enumeration of polynomials generates an enumeration of their roots; when a polynomial has more than one real root, we enumerate its roots in numerical order.

To obtain an α_n and β_n such that $\beta_n - \alpha_n \leq 1/k$, we first enumerate those roots of polynomials of size $2k$ or less that are in the interval $(0, 1)$. Applying the procedure in Cantor's proof to this enumeration, we obtain the finite sequence $\alpha_1, \beta_1, \alpha_2, \beta_2, \ldots, \alpha_n, \beta_n$. Now if $\beta_n - \alpha_n > 1/k$, then we would have $\alpha_n < j/(2k) < (j + 1)/(2k) < \beta_n$ for some j between 1 and $2k - 2$. Since $j/(2k)$ and $(j + 1)/(2k)$ are roots of polynomials of size $2k$, we have found two roots of our enumeration between α_n and β_n. But Cantor's procedure allows at most one of these roots to be between α_n and β_n. Hence, we must have $\beta_n - \alpha_n \leq 1/k$. Since there are $(4k + 1)^{2k+1} - (4k + 1)$ polynomials of size $2k$ or less, and since each one has at most $2k$ roots, we need to examine at most $2k[(4k + 1)^{2k+1} - (4k + 1)]$ algebraic reals in order to approximate our transcendental η to within $1/k$.

This simple argument produces a poor bound. Figure 1 shows that we do not need to enumerate polynomials of size 200 to obtain approximations differing by less than $1/100$. Nevertheless, our computer program generates digits inefficiently

—asymptotically, it takes at least $O(2^{\sqrt[3]{n}})$ steps to generate n digits of our transcendental number [**18**]. Any program requiring this many steps is regarded as inefficient by computer scientists [**16**, p. 6–9].

Cantor's proof leads to a computer program that generates a transcendental, but a program is not necessary for understanding his article. The constructive nature of Cantor's article is clear from the wording and proof of Theorem 2. This theorem separates the constructive part of his article from the proof-by-contradiction needed to establish the uncountability of the set of reals. Since we will be referring to Theorem 2 throughout our article, we shall give it a name—*Cantor's theorem on real sequences*.

We are far from the first to point out that Cantor's article is constructive. In 1930, Fraenkel stated that the method in this article is "a method that incidentally, contrary to a widespread interpretation, is fundamentally constructive and not merely existential" [**15**, p. 237].

Exercise. The sequence $1/2$, $1/3$, $2/3$, $1/4$, $2/4$, $3/4, \dots$. contains all the rationals belonging to $(0, 1)$. Apply the algorithm in Cantor's proof to this sequence to generate the digits of an irrational. If you are using pencil and paper, just compute α_1, β_1, α_2, and β_2.

3. CANTOR'S DIAGONAL PROOF. We now turn to Cantor's 1891 article [**9**], which contains his well-known diagonal proof. Cantor begins by discussing his 1874 article. He points out that it contains a proof of the theorem: There are infinite sets that cannot be put into one-to-one correspondence with the set of positive integers. Then he asserts that this theorem has a much simpler proof than the one given in 1874. His new proof uses the set M of elements of the form $E = (x_1, x_2, \dots, x_\nu, \dots)$, where each x_ν is either m or w. Cantor states that M is uncountable, and notes that this result is implied by the following theorem:

> *If $E_1, E_2, \dots, E_\nu, \dots$ is any simply infinite sequence of elements of the set M, then there is always an element E_0 of M which corresponds to no E_ν.*

Cantor proves his theorem by using the diagonal method to construct E_0. Note that, once again, Cantor states a theorem that separates the constructive content of his work from the proof-by-contradiction needed to establish uncountability.

By introducing sequences of abstract symbols, Cantor shows that the phenomenon of uncountability does not depend on properties of the real numbers, such as the existence of limits for bounded increasing (or decreasing) sequences of reals. Thus, Cantor shows that uncountability is a fundamental phenomenon of set theory. Also, his 1874 theorem on real sequences follows easily from his new theorem. Take any sequence of real numbers and expand its members into their binary representations. (A real of the form $m/2^n$ must be expanded into both of its binary representations.) This gives us a sequence of binary representations. Apply Cantor's new theorem to obtain the binary representation of a real number that does not belong to the original sequence.

Cantor's diagonal method is simpler than his earlier nesting method, and it generates transcendentals much more efficiently. The diagonal method can generate n digits of a transcendental in $O\ (n^2 \log^2 n \log\log n)$ steps [**18**]. Algorithms requiring less than $O(n^3)$ steps are considered practical by computer scientists [**16**, p. 9].

Figure 2 contains output from a computer program that uses Cantor's diagonal method to generate the digits of a transcendental in $(0, 1)$. Our program generates the same ω_n sequence as our program in Section 2. It uses this sequence to generate the digits of a diagonal number as follows: Let d be the nth digit of ω_n. Our program sets the nth digit of our diagonal number to $d + 1$ unless d is 9; in this case, it sets the nth digit to 0. Cantor's diagonal argument guarantees that the decimal representation of our diagonal number differs from the representations we use for the algebraic reals. But as Figure 2 shows, our program does not generate both representations of fractions such as $1/2$. Hence, we cannot conclude that our diagonal number is transcendental until we show that it differs from all fractions having two decimal representations. Our diagonal number does differ from these fractions because its decimal expansion contains infinitely many 2's—the decimal expansions of $1/9, 1/90, 1/900, \ldots$ generate 2's on the diagonal.

Note that our diagonal number can be written as:

$$\sum_{n=1}^{\infty} \frac{\operatorname{rem}(\lfloor 10^n \cdot \omega_n \rfloor + 1, 10)}{10^n}$$

where ω_n is the nth member of our sequence of algebraic reals; $\operatorname{rem}(m, n)$ is the remainder left after dividing m by n; and $\lfloor x \rfloor$ is the *floor* of x (the largest integer equal to or less than x).

Exercise. Write the rationals in $(0, 1)$ as a sequence: $1/2$, $1/3$, $2/3$, $1/4$, $2/4$, $3/4, \ldots$. By applying the diagonal method to this sequence, generate an irrational number in $(0, 1)$. Compute the first 10 digits of this number, compute the 25th

Algebraic Real	Associated Polynomial	Approximation to Transcendental
$\omega_1 = .5$	$2x - 1$	.6
$\omega_2 = .61\ldots$	$x^2 + x - 1$	.62
$\omega_3 = .732\ldots$	$x^2 + 2x - 2$	.623
$\omega_4 = .4142\ldots$	$x^2 + 2x - 1$	.6233
$\omega_5 = .70710\ldots$	$2x^2 - 1$	.62331
$\omega_6 = .780776\ldots$	$2x^2 + x - 2$	.623317
$\omega_7 = .3660254\ldots$	$2x^2 + 2x - 1$	.6233175
$\omega_8 = .66666666\ldots$	$3x - 2$	.62331757
$\omega_9 = .333333333\ldots$	$3x - 1$	.623317574
$\omega_{10} = .3819660112\ldots$	$x^2 - 3x + 1$	.6233175743
$\omega_{11} = .79128784747\ldots$	$x^2 + 3x - 3$	.62331757438
$\omega_{12} = .561552812808\ldots$	$x^2 + 3x - 2$	.623317574389
$\omega_{13} = .3027756377319\ldots$	$x^2 + 3x - 1$	.6233175743890
$\omega_{14} = .82287565553229\ldots$	$2x^2 + 2x - 3$	.62331757438900
$\omega_{15} = .686140661634507\ldots$	$2x^2 + 3x - 3$	.623317574389008

Figure 2. Generating a transcendental using Cantor's diagonal method.

digit. Verify that this number can be written as:

$$\sum_{n=2}^{\infty} \sum_{m=1}^{n-1} \frac{\operatorname{rem}\left(\left\lfloor 10^{f(m,n)} \cdot \frac{m}{n} \right\rfloor + 1, 10\right)}{10^{f(m,n)}}$$

where $f(m,n) = (n-2)(n-1)/2 + m$.

4. ALL TRANSCENDENTALS LIVE ON DIAGONALS. As we have seen, Cantor's diagonal method does construct transcendentals—but which ones? To answer this question, we first observe that the digits generated by the diagonal method depend on the enumeration of algebraic reals we use. Different enumerations usually lead to different transcendentals. Since there are $2^{\aleph_0}$ transcendentals and $2^{\aleph_0}$ enumerations of the algebraic reals, perhaps all transcendentals live on diagonals. In a precise sense, they do.

Before we can state our theorem about diagonals and transcendentals, we need some definitions. Let $b(n)$ denote the nth digit of b, where b is the binary representation of a real number. We define the *diagonal number* of the sequence $b_1, b_2, b_3, \ldots$ of binary representations to be the real number whose binary representation d is obtained by the following rule: $d(n) = 0$ if $b_n(n) = 1$, and $d(n) = 1$ if $b_n(n) = 0$. (With binary representations, there is only one way to change a digit, so there is only one diagonal number associated with a sequence. To work with other representations, we would have to talk about the *diagonal numbers* of a sequence.) We say that a sequence *consists of all the binary representations of algebraic reals* if it contains all the binary representations of the algebraic reals (including both representations of the fractions $m/2^n$) and if it does not contain any representations of non-algebraic reals. Such a sequence may contain the same representation more than once. Using the above definitions, we can express the relationship between transcendentals and diagonal numbers:

Theorem 3. *A real number in the interval* $(0,1)$ *is transcendental if and only if it is the diagonal number of a sequence that consists of all the binary representations of algebraic reals in* $(0,1)$.

Proof: By Cantor's diagonal argument, the diagonal number of such a sequence is transcendental.

Now assume that t is transcendental. Let a_n be any sequence consisting of all the binary representations of algebraic reals in the interval $(0,1)$. We will define a sequence b_n that is a permutation of the sequence a_n and that generates t as its diagonal number.

Throughout our proof, we use the same notation to denote a real number and its binary representation. We start by finding the first a_k such that $a_k(1) \neq t(1)$. Our search is bounded by the binary representations of $1/2$ since one of these representations starts with 1 and the other starts with 0. After finding our a_k, we mark it as used and set $b_1 = a_k$. Now assume that we have found $b_1, b_2, \ldots, b_{n-1}$. To obtain a suitable b_n, we look for the first unused a_k such that $a_k(n) \neq t(n)$. If $t(n) = 0$, then our search is bounded by the binary representations of $1/2^n + 1/2^{n+i}$ where $i = 1, \ldots, n$. The representations of these numbers have a 1 in their nth place and at least one of them is unused. Similarly, if $t(n) = 1$, then our search is bounded by the binary representations of $1/2^{n+i}$ where $i = 1, \ldots, n$. After finding our a_k, we mark it as used and set $b_n = a_k$.

To complete our proof, we must show that the sequence b_n is a permutation of the sequence a_n. Assume that some a_n was not used, and let a_k be the unused one with the least index. Now each a_i, for $i < k$, was used to define a b_j. Let N be greater than the indices of these b_j's. (If $k = 1$, then there are no such b_j's so we let N be 1.) By our definition of the sequence b_n, the only way for a_k to stay unused is for the equality $a_k(n) = t(n)$ to hold for all $n \geq N$. Hence, $t - a_k$ is rational. Since t is transcendental, a_k must also be transcendental—but this contradicts the fact that a_k is an algebraic real. Thus, the sequence b_n is a permutation of the sequence a_n.

If we apply the method in our proof to a transcendental t whose digits are computable, then we can compute a sequence of algebraic reals whose diagonal number is t. For example, we have written a computer program that uses the binary representation of $1/e$ to generate a sequence of algebraic reals whose diagonal number is $1/e$. Output from this program is given in Figure 3. The sequence generated by our program is a permutation of a sequence ω_n that consists of all the binary representations of algebraic reals in the interval $(0, 1)$. This ω_n sequence is similar to the ω_n sequence of our previous programs—the key difference is that our new sequence consists of representations rather than numbers. Since our new sequence contains both binary representations of the fractions $m/2^n$, its numbering differs from that of our previous sequence. For example, in Figure 3, both ω_1 and ω_2 are binary representations of $1/2$.

Approximation to $1/e$	Binary Representation of Algebraic Real	Associated Polynomial
$.\mathbf{0}$	$\omega_2 = .\mathbf{1}$	$2x - 1$
$.0\mathbf{1}$	$\omega_3 = .1\mathbf{0}\ldots$	$x^2 + x - 1$
$.01\mathbf{0}$	$\omega_1 = .01\mathbf{1}\ldots$	$2x - 1$
$.010\mathbf{1}$	$\omega_5 = .011\mathbf{0}\ldots$	$x^2 + 2x - 1$
$.0101\mathbf{1}$	$\omega_6 = .1011\mathbf{0}\ldots$	$2x^2 - 1$
$.01011\mathbf{1}$	$\omega_4 = .10111\mathbf{0}\ldots$	$x^2 + 2x - 2$
$.010111\mathbf{1}$	$\omega_8 = .010111\mathbf{0}\ldots$	$2x^2 + 2x - 1$
$.0101111\mathbf{0}$	$\omega_7 = .1100011\mathbf{1}\ldots$	$2x^2 + x - 2$
$.01011110\mathbf{0}$	$\omega_9 = .10101010\mathbf{1}\ldots$	$3x - 2$
$.010111100\mathbf{0}$	$\omega_{10} = .010101010\mathbf{1}\ldots$	$3x - 1$

Figure 3. Generating a sequence whose diagonal number is $1/e$.

Exercise. Construct a sequence consisting of all the binary representations of rationals in $(0, 1)$ by expanding the rationals $1/2, 1/3, 2/3, 1/4, 2/4, 3/4, \ldots$ into their binary representations. Now use the algorithm in the proof of Theorem 3 to generate the first 10 members of a sequence of binary representations whose diagonal number is the irrational $\sqrt{2} - 1$. (The binary representation of $\sqrt{2} - 1$ is $0.0110101000\ldots$.)

5. CANTOR'S UNPUBLISHED PROOF. In Section 2, we saw that Cantor's ideas can be used to write either a direct (constructive) proof or an indirect (non-constructive) proof of the existence of transcendentals. We now investigate whether Cantor knew that his ideas could produce such different proofs.

The thinking that led to Cantor's 1874 article can be found in the correspondence between Cantor and Dedekind [**11**, p. 187–191; **28**, p. 12–16]. We start with Cantor's letter of November 29, 1873, in which he asks Dedekind the following question:

> Take the collection of all positive whole numbers n and denote it by (n); further, imagine the collection of all positive real numbers x and denote it by (x); the question is simply whether (n) and (x) can be corresponded so that each individual of one collection corresponds to one and only one individual of the other.

Cantor says that at first glance it appears that no such correspondence could exist —after all, one collection is discrete and the other continuous. But then he brings out the subtlety of his question by stating that it is easy to construct a one-to-one correspondence between the collection of positive integers and the collection of rational numbers. Cantor also states that a one-to-one correspondence can be constructed between the collection of positive integers and general collections of the form $(a_{n_1, n_2, \ldots, n_\nu})$, where the indices $n_1, n_2, \ldots, n_\nu$ and ν are positive integers.

Dedekind replies that he is unable to answer the question, but he does give Cantor a one-to-one correspondence between the collection of algebraic numbers and the collection of positive integers. Dedekind also advises Cantor not to waste too much time on his question because it has no "particular practical interest."

In his next letter, dated December 2nd, Cantor acknowledges Dedekind's advice but points out that his question is of interest: "It would be nice if it could be answered; for example, provided that it were answered *no*, one would have a new proof of Liouville's theorem that there are transcendental numbers."

Cantor's letter of December 7th contains the result he is seeking. His proof starts:

> Suppose that the positive numbers $\omega < 1$ can be broken up into the sequence:
>
> $(I) \qquad \omega_1, \omega_2, \ldots, \omega_n, \ldots$

After an involved argument, Cantor obtains a contradiction.

In his next letter, dated December 9th, Cantor outlines the proof he will publish:

> I show directly that if I start with a sequence
>
> $(I) \qquad \omega_1, \omega_2, \ldots, \omega_n, \ldots$
>
> I can determine, in *every* given interval $(\alpha..\beta)$, a number η that is not included in (I). Hence, it follows immediately that the collection (x) cannot correspond one-to-one with the collection (n)...

Taken together, Cantor's letters of December 2nd and 7th provide an indirect proof of the existence of transcendentals. But his letter of December 9th contains his theorem on real sequences, which provides a direct construction of transcendentals.

6. WHY IS CANTOR'S ARTICLE MISINTERPRETED? Cantor's correspondence with Dedekind, which contains his indirect existence proof, was not published until 1937 [**17**, p. 104]. By then, other mathematicians had rediscovered the proof. Klein outlined it in 1894 [**21**, p. 51]. Since Klein (as far as we know) published the proof first, we will call it *Klein's proof*. In 1907, Osgood presented Klein's proof, but called it "Cantor's proof for the existence of non-algebraic numbers"

[**29**, p. 159–160].* In 1921, Perron presented Klein's proof, attributed it to Cantor, and then critiqued it [**30**, p. 161–162]:

> ...Cantor's proof for the existence of transcendental numbers has, along with all its simplicity and elegance, the great disadvantage that it is only an existence proof; it does not enable us to actually specify even a single transcendental number. Free from this disadvantage is another—in fact, the oldest—existence proof due to Liouville...

Perron's critique is similar to the one given by Kac and Ulam (see Section 1). So this view of Cantor's work has been around for many years. Why do some mathematicians misinterpret Cantor's article? Undoubtedly, there are a variety of reasons. We will only discuss how these misinterpretations are encouraged by some commonly-held views about mathematics and its history.

One such view states that Cantor's set theory was initially attacked by many mathematicians of his time. The problem with this view is that it fails to distinguish the parts of Cantor's work that were attacked, from those that were not. For example, consider the following statement by Birkhoff and MacLane [**1**, p. 436–437]:

> Cantor's argument for this result ["Not every real number is algebraic"] was at first rejected by many mathematicians, since it did not exhibit any specific transcendental number.

If Cantor's argument was rejected, then it would be reasonable to suspect that his argument is non-constructive and that this was the reason for its rejection. However, we have found no evidence indicating that Cantor's argument was rejected. Kronecker—the mathematician most likely to reject it—had a chance to, but did not.

Cantor sent the article containing his existence proof to *Crelle's Journal* (*Journal für die reine und angewandte Mathematik*), even though he knew that Kronecker, as one of the journal's editors, could reject or delay the article. Previously, Kronecker had delayed the publication of an article written by Heine, one of Cantor's colleagues. In fact, Kronecker had even tried to persuade Heine to withdraw his article [**12**, p. 67 and p. 308–309]. Cantor's experience was different —his article was printed quickly. Apparently, Kronecker found it to be no worse (from his point of view) than other articles appearing in *Crelle's Journal*.

Cantor's article does contain revolutionary ideas, but Cantor rephrases these ideas using terminology familiar to his contemporaries. For example, he introduces the concept of a collection of reals corresponding one-to-one with the collection of positive integers, and then he provides an equivalent formulation—namely, that such a collection of reals can be written as a sequence. Also, he states the two theorems of his article in terms of sequences rather than one-to-one correspondences. Finally, Cantor incorporates a constructive idea into his article—careful reading reveals that his theorems only deal with sequences that are ordered by a "law" (see [**3**, p. 260; **5**, p. 308] and our first Cantor quotation in Section 2). Cantor may have inserted this restriction to avoid problems with Kronecker. Kronecker required that a sequence or series be generated by an arithmetic rule—in fact, Kronecker probably objected to Heine's article because it dealt with *arbitrary* trigonometric series [**14**, p. 71].

---

*Klein never attributed his proof to Cantor. In 1895, Klein gave an exposition of Cantor's 1874 article in which he replaced Cantor's old nesting method with the newer diagonal method. Klein called the resulting constructive proof "Cantor's proof" [**22**, p. 49–54].

In his next article [4, 6], which was published in 1878, Cantor introduces concepts that apply to *all* sets—these concepts cannot be rephrased into the mathematical terminology of his time. Cantor begins this article by introducing the notion of a one-to-one correspondence between two arbitrary sets, finite or infinite. Next he defines the ordering that this correspondence induces—that is, what it means for one set to be of lesser, equal, or greater *power* (*cardinality*) than another. Cantor devotes most of this article to proving that for every positive integer n, the set of all n-tuples of reals can be put into one-to-one correspondence with the set of reals.

Once again, Cantor submitted his work to *Crelle's Journal*—but this time, publication was delayed. Cantor blamed Kronecker for the delay and never sent another article to *Crelle's Journal* [12, p. 69–70; 17, p. 111–113]. We have no record of why Kronecker had difficulty with Cantor's article. However, to understand this article, one must work with countably infinite and uncountably infinite sets. Kronecker, with his constructive philosophy, could never accept Cantor's reasoning.

As Cantor developed his ideas about infinity further, more mathematicians began to criticize his work. For example, Cantor presented his theory of transfinite ordinal numbers in his 1883 monograph *Foundations of a General Theory of Sets* [7, 8]. After reading Cantor's monograph, Poincaré made the following comment [13, p. 278; 27, p. 95–96]:

> ... these numbers in the second, and especially in the third, number-class have the appearance of being form without substance, something repugnant to the French mind.

(Cantor's first number class consists of the natural numbers. His second number class consists of the countable ordinals, which are the ordinals representing the well-orderings of the first number class. His third number class consists of the ordinals that represent the well-orderings of the second number class.)

So the popular view of the history of set theory needs to be refined. Criticism of Cantor's theory did not begin with the publication of his 1874 article. It began with his 1878 article, which contains arguments that require the use of infinite sets. Criticism increased as Cantor introduced new concepts involving the infinite.

A commonly-held view about existence proofs also needs examining— namely, the view stating that most non-constructive existence proofs are simpler than their constructive counterparts. Most non-constructive proofs are simpler. However, we must recognize those cases in which a construction yields the simplest proof. For example, Perron (see quotation above) mentions the "simplicity and elegance" of the non-constructive existence proof that we call Klein's proof. Comparing Klein's proof to Liouville's constructive proof, it is tempting to conclude that the former is simpler because of its non-constructive nature. However, Cantor's constructive approach yields an even simpler proof.

Klein's proof requires that we first prove that the set of reals is uncountable. How do we prove this? By assuming that there is a sequence that enumerates the reals, applying the diagonal method to this sequence, and obtaining a contradiction. Dissecting this proof, we find that it rests upon two facts:

(1) If we apply the diagonal method to a sequence of reals, then we obtain a real not in the sequence.
(2) If we assume that the reals can be enumerated by a sequence and obtain a contradiction from this assumption, then no such enumerating sequence exists.

Now if we apply the diagonal method to a sequence containing all the algebraic reals, we only need fact (1) to guarantee that we have constructed a transcendental. So Cantor's constructive approach is simpler than the non-constructive one (and it provides excellent preparation for the uncountability proof).

There may be other views about mathematics and its history that lead some mathematicians to misinterpret Cantor's article. We encourage our readers to explore this subject further.

7. CONCLUSION. We started this article with two conflicting statements about Cantor's proof of the existence of transcendentals. We have seen that these statements are talking about two different proofs of the existence of transcendentals, a constructive proof that goes back to Cantor's original articles and a non-constructive proof that appeared later.*

We advocate teaching either both proofs or just the constructive one. As mentioned in our last section, Cantor's constructive approach is simpler than the non-constructive one. Also, while discussing Cantor's approach, we can give him the credit he deserves for presenting his work so constructively.

APPENDIX. Figures 1 and 2 (see Sections 2 and 3) contain output from computer programs that use Cantor's methods. Both of these programs generate a sequence ω_n of algebraic reals. We now define this sequence precisely so that interested readers may verify our results.

As discussed in Section 2, we enumerate the polynomials with integer coefficients by their size and we use this enumeration to generate an enumeration of the algebraic reals. We only need to enumerate irreducible polynomials; but since Cantor's methods do work with sequences containing duplicates, we did not bother to check for irreducibility. (For those who are interested in testing for irreducibility, see [**23**, p. 431–434].) However, to simplify our calculations and avoid many duplicate roots, we decided to enumerate polynomials $p(x)$ with the following properties:

(1) $p(x)$ has a root in the interval $(0, 1)$.
(2) The coefficients of $p(x)$ have no common factor greater than one, and the leading coefficient is positive.
(3) Either $p(x)$ is linear or it has no linear factors.
(4) $p(x)$ and its derivative $p'(x)$ have no common roots.
(5) $p(x)$ and its second derivative $p''(x)$ have no common roots.

The first condition simplifies our calculations. The second and third conditions eliminate many duplicate roots.

The last two conditions are needed by Newton's method. As Figure 1 shows (see Section 2), Cantor's 1874 nesting method generates digits very slowly. So for this method, we only used Sturm's theorem and Horner's method to approximate roots [**31**, p. 138–156]. But the diagonal method generates digits much faster than the nesting method. To generate a large number of these digits, we need an efficient approximation algorithm, such as Newton's method. Now to guarantee that Newton's method does converge to a root of $p(x)$, we first use Sturm's theorem and Horner's method to isolate our root to an interval where $p'(x)$ and $p''(x)$ do

*That is, appeared in print later—our study of Cantor's correspondence (see Section 5) shows that he knew both proofs.

not vanish. Since the polynomials in our enumeration do not share roots with their first or second derivatives, our root does belong to such an interval. After isolating our root, we set our initial approximation to an endpoint of this interval, and then we use Newton's method to generate better approximations (see [**31**, p. 174–177] for details—such as, which endpoint to choose for the initial approximation).

ACKNOWLEDGMENTS. I am very grateful to Michael Broido, Joseph Dauben, William Dunham, Stephen Maurer, Edward Sandifer, James Tattersall, and the referees for reading and commenting on drafts of this article. Their suggestions have led to numerous improvements. However, I am solely responsible for all opinions as well as any remaining errors. I am also responsible for the accuracy of all translations. Here I was aided by the French translations of Cantor's articles and letters [**5**, **11**] as well as by Dauben's English translations [**12**]. My historical research was aided by Dauben's excellent biography of Cantor [**12**]. The photograph of Cantor on the cover is through the courtesy of the late Wilhelm Stahl and Ivor Grattan-Guinness. It shows Cantor in his middle years. Comparison with the photographs in [**26**, p. 530–531] suggests that it was taken between 1880 and 1894.

REFERENCES

1. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 4th ed., Macmillan, New York, 1977. Earlier editions: 1941, 1953, 1965.
2. E. Bishop and D. Bridges, *Constructive Analysis*, Springer-Verlag, Berlin, 1985.
3. G. Cantor, Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen, *J. Reine Angew. Math.* **77** (1874), 258–262. Reprint: [**10**, p. 115–118].
4. G. Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, *J. Reine Angew. Math.* 84 (1878), 242–258. Reprint: [**10**, p. 119–133].
5. G. Cantor, Sur une propriété du système de tous les nombres algébriques réels, *Acta Math.* 2 (1883), 305–310. (French translation of [**3**].)
6. G. Cantor, Une contribution à la théorie des ensembles, *Acta Math.* 2 (1883), 311–328. (French translation of [**4**].)
7. G. Cantor, *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*, Teubner, Leipzig, 1883. Reprint: [**10**, p. 165–208].
8. G. Cantor, Fondaments d'une théorie générale des ensembles, *Acta Math.* 2 (1883), 381–408. (French translation of the mathematical part of [**7**]; the philosophical part was not translated.)
9. G. Cantor, Über eine elementare Frage der Mannigfaltigkeitslehre, *Jahresbericht der Deutschen Mathematiker-Vereinigung* 1 (1891), 75–78. Reprint: [**10**, p. 278–280].
10. G. Cantor, *Gessammelte Abhandlungen mathematischen und philosophischen Inhalts*, ed. E. Zermelo, Springer, Berlin, 1932. Reprint: Hildesheim, Olms, 1966.
11. J. Cavaillès, *Philosophie mathématique*, Hermann, Paris, 1962.
12. J. W. Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, 1978.
13. P. Dugac, *Lettres de Charles Hermite à Gösta Mittag-Leffler (1874–1883)*, Cahiers du Séminaire d'Histoire des Mathématiques 5 (1984), 49–285.
14. H. M. Edwards, Kronecker's Views on the Foundations of Mathematics, *The History of Modern Mathematics, Vol. 1*, (edited by D. E. Rowe and J. McCleary), Academic Press, Boston, 1989, 67–77.
15. A. Fraenkel, Georg Cantor, *Jahresbericht der Deutschen Mathematiker-Vereinigung* 39 (1930), 189–266.
16. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, San Francisco, 1979.
17. I. Grattan-Guinness, The rediscovery of the Cantor-Dedekind correspondence, *Jahresbericht der Deutschen Mathematiker-Vereinigung* 76 (1974), 104–139.
18. R. Gray, The computational complexity of Cantor's transcendentals, (in preparation).
19. I. N. Herstein and I. Kaplansky, *Matters Mathematical*, Harper & Row, New York, 1974.
20. M. Kac and S. M. Ulam, *Mathematics and Logic*, Frederick A. Praeger, New York, 1968.
21. F. Klein, *Lectures on Mathematics*, Macmillan, New York, 1894.
22. F. Klein, *Vorträge über Ausgewählte Fragen der Elementargeometrie*, Teubner, Leipzig, 1895. Translated by W. W. Beman and D. E. Smith as *Famous Problems of Elementary Geometry*, Ginn, Boston, 1897; reprinted by Chelsea Publishing Company, New York, 1955. Page references are to English translation.

23. D. Knuth, *The Art of Computer Programming Vol. 2*, 2nd ed., Addison-Wesley, Reading, Mass., 1981.
24. J. Liouville, Remarques sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques, *Comp. Rend. Acad. Sci. Paris* 18 (1844), 883–885.
25. J. Liouville, Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques, *J. Math. Pures Appl.* 16 (1851), 133–142.
26. H. Meschkowski and W. Nilson, eds., *Georg Cantor Briefe*, Springer-Verlag, Berlin, 1991.
27. G. H. Moore, Towards A History of Cantor's Continuum Problem, *The History of Modern Mathematics, Vol. 1*, (edited by D. E. Rowe and J. McCleary), Academic Press, Boston, 1989, 79–121.
28. E. Noether and J. Cavaillès, eds., *Briefwechsel Cantor-Dedekind*, Hermann, Paris, 1937.
29. W. F. Osgood, *Lehrbuch der Funktionentheorie*, Teubner, Leipzig, 1907. Reprint of 1928 edition: Chelsea, New York, 1965. (In reprint, Cantor's work is mentioned on p. 203–204.)
30. O. Perron, *Irrationalzahlen*, de Gruyter, Berlin, 1921. Reprint of 1929 edition: Chelsea, New York, 1951. (In reprint, Cantor's work is mentioned on p. 174.)
31. J. V. Uspensky, *Theory of Equations*, McGraw-Hill, New York, 1948.

5 Joseph Road
Framingham, MA 01701

Added in Proof. The quotation of Poincaré in Section 6 is out of context. It comes from a letter explaining why French mathematicians may not appreciate a proposed translation of Cantor's 1883 monograph—a monograph that Poincaré twice calls "beautiful" [**13**, p. 278]. Poincaré warns that, unless they are given concrete examples, mathematicians unfamiliar with Cantor's previous work may find his ordinals to be "form without substance."

A quotation that accurately reflects some of the criticism of the time can be found in an 1883 letter from Hermite to Mittag-Leffler [**13**, p. 209; **27**, p. 96]:

> The impression that Cantor's memoirs makes on us is distressing. Reading them seems, to all of us, to be a genuine torture .... While recognizing that he has opened up a new field of research, none of us is tempted to pursue it. For us it has been impossible to find, among the results that can be understood, a single one having *current interest*. The correspondence between the points of a line and a surface leaves us absolutely indifferent and we think that this result, as long as no one has deduced anything from it, stems from such arbitrary methods that the author would have done bettter to withhold it and wait.

Hermite's "us" includes Appell, who read a draft of the letter, and probably Picard, who was also critical of Cantor's recent work [**13**, p. 212]. The criticism at the end of the quotation is directed at the major result of Cantor's 1878 article.