

Real Numbers and Compression

This short document tries to convey the strange nature of real numbers and the reasons why many mathematicians feel that most real numbers are unreal; that in some sense they do not exist. Compression is mentioned only once, toward the end.

We feel that we “understand” integers intuitively (because we can count one cow, two cows, etc.), but when we think deeply about real numbers we realize that they have such unexpected properties and exhibit such unintuitive behavior, that there is no hope of ever “understanding” them and there are also reasons to claim that most of them do not exist (because they cannot be computed and cannot be named).

The real numbers can be divided into the sets of rational and irrational. A rational number can be represented as the ratio of two integers, whereas an irrational number cannot be represented in this way. The ancient Greeks already knew that $\sqrt{2}$ is irrational. The real numbers can also be divided into algebraic and transcendental numbers. The former is the set of all the reals that are solutions of algebraic equations.

We know many integers (0, 1, 7, 10, and 10^{100} immediately come to mind). We are also familiar with a few irrational numbers ($\sqrt{2}$, e , and π are common examples), so we intuitively feel that most real numbers must be rational and the irrationals are a small minority. Similarly, it is easy to believe that most reals are algebraic and that transcendental numbers are rare. However, set theory, the creation, in the 1870s, of Georg Cantor, shows that there are different kinds of infinities, that the reals constitute a greater infinity than the integers (the integers are said to be countable, while the reals are uncountable), that the rational numbers are countable, while the irrationals are uncountable, and similarly, that the algebraic numbers are countable, while the transcendentals are uncountable; completely counterintuitive notions.

Today, we believe in the existence of atoms. If we start with a chunk of matter, cut it into pieces, cut any piece into smaller pieces, and continue in this way, we eventually arrive at individual atoms or even their constituents. Also, the more massive an object is, the more atoms it consists of. The real numbers, however, are very different. They can be represented as points along an infinitely long number line, but they are everywhere dense on this line. Any segment on the number line, as short as we can imagine, contains an (uncountable) infinity of real numbers. We cannot arrive at a segment containing just one number by repeatedly segmenting and producing shorter and shorter segments. On the other hand, we can prove that two segments of different lengths contain the same infinity of reals.

Imagine a straight segment of unit length and denote its midpoint by M . Now assign to each point Q on the segment a coordinate d_Q that is its distance from M . Points to the right of M have positive coordinates up to $1/2$, while points to the left of M have negative coordinates down to $-1/2$. Now imagine an infinitely-long straight line. It does not have a midpoint, so we select a point N at random as our midpoint. Each point P on the line is assigned as its coordinate its distance d_P from N . Clearly, these distances vary from $-\infty$ to $+\infty$. We now show that our short segment and long line have the same number of points (the same infinity) by associating a point Q on the segment for each point P on the line. We do this with the simple function $d_Q = \arctan(d_P)/\pi$.

For those who forgot their school trigonometry, the arctan function is the inverse of tan (or tangent). This is a multi-valued function whose main branch varies from $\arctan(-\infty) = -\pi/2$ to $\arctan(0) = 0$ to $\arctan(\infty) = \pi/2$. Thus $\arctan(d_p)/\pi$ varies from $-1/2$ to $+1/2$ and covers the coordinates d_Q of all the points Q on the 1-unit-long segment.

This simple function associates a point Q on the segment with every point P on the infinite line, thereby showing that the segment and the line have the same infinity of points, an unintuitive result. This result is true for any segment, regardless of its length, which implies that any interval of real numbers, even the shortest imaginable, contains the same infinity of reals as the interval $[-\infty, +\infty]$.

Even more surprises await the reader. It is easy to show that the (infinite) number of two-dimensional points equals the (infinite) number of reals. All that we have to do is find a function that will map each two-dimensional point (x, y) to a real number R such that different points will map to different reals.

A two-dimensional point is a pair of coordinates (real numbers), such as 6.9226543 and 4.0358165. A simple way to map such a pair to a real number is to interleave their digits. Thus, the two numbers above become the real number 0.6490232568514635. All the digits of the two coordinates are used, each is used once, none is duplicated, and none is deleted. It is clear that a different point would be mapped to a different real number, which shows that our mapping is one-to-one.

The next surprise involves the well-known Cantor dust. This is a set of real numbers that is constructed iteratively, in an infinite refinement process. We describe the modern version, known as the Cantor ternary set. Start with the closed interval $[0, 1]$ and delete the middle open third $(1/3, 2/3)$. This leaves the two closed intervals $[0, 1/3]$ and $[2/3, 1]$. Delete the middle open third of each, and continue in this way indefinitely. After each step, the number of intervals left in the set is doubled, but each is $1/3$ of the size of the intervals of the preceding step. In the limit, the number of intervals left in the set becomes infinite, and each is infinitely short, thereby justifying the name dust.

Cantor showed that even this insignificant dust is uncountable, because it can be put into a one-to-one correspondence with all the real numbers in the original interval $[0, 1]$. This mapping is especially easy to understand if we consider the ternary representations of the numbers that constitute the dust. Later in this document we show that the integer 1 can be written as the repeating fraction $0.999\dots$. Similarly, the value $1/3 = 0.1_3$ can be written as the repeating ternary fraction $0.0222\dots_3$ and $2/3 = 0.2_3$ can be written as $0.1222\dots_3$. The first time a middle third is removed, it contains numbers of the form $0.1xxx\dots_3$, where $xxx\dots_3$ is between $0000\dots_3$ and $2222\dots_3$. Thus, the numbers remaining after the first step are of one of the following forms:

- Numbers of the form $0.0xxxx\dots_3$.
- $1/3 = 0.1_3 = 0.02222\dots_3$.
- $2/3 = 0.12222\dots_3 = 0.2_3$.
- Numbers of the form $0.2xxxx\dots_3$.

It does not take long to realize that these remaining numbers have either 0 or 2 as the first (ternary) digit after the ternary point. All the numbers of the form $0.1xxx\dots_3$

have been removed in the first step. Similarly, the second step retains these numbers that have 0 or 2 in the second position after the ternary point and removes those numbers that have a 1 at that position. Thus, the final dust consists of those ternary numbers in the closed interval $[0, 1]$ that do not have a 1 in their ternary representation, such as, e.g., 0.02220200200002_3 .

Now comes the mapping. Given a number in the Cantor dust, first change each of its 2's to a 1, then think of it as a binary number. Thus, the ternary number 0.02220200200002 is mapped to the binary number 0.01110100100001 . Every member of the dust is mapped to some number in the interval $[0, 1]$ and every such number is mapped to a member of the dust. Ingenious!

We are also familiar with the concepts of successor and predecessor. An integer N has both a successor $N + 1$ and a predecessor $N - 1$. Cantor has shown that the rational numbers are countable; each can be associated with an integer (i.e., each rational can be assigned an integer tag or subscript). Thus, each rational number can be said to have a successor and a predecessor. The real numbers, again, are different. Given a real number a , we cannot point to its successor. If we find another real number b that may be the successor of a , then there is always another number, namely $(a + b)/2$, that is located between a and b and is thus closer to a than b is. We therefore say that a real number does not have a successor or a predecessor; it does not have any immediate neighbors. This is why the term continuum is used to refer to the set of reals. Every point P on the number line has a real number that corresponds to it (it equals the distance of P from point 0 with infinite precision), but it is impossible to move from P to the “next” point. Thus, the real number line features a duality of attributes. It is both discrete (because we can land at any point) and continuous (because a point does not have immediate neighbors). We cannot imagine any collection of points, numbers, or any other objects that are everywhere (extremely) dense but do not feature a predecessor/successor relation. The real numbers are therefore very counterintuitive.

The concept of a continuum is baffling, so here is an attempt to illuminate it. Imagine the interval $[0, 1]$. This is a closed interval that contains all the real numbers between 0 and 1, including the endpoints 0 and 1. Now remove the two endpoints to obtain the open interval $(0, 1)$. Starting at the midpoint 0.5, move steadily to the right, encountering larger and larger numbers. We meet numbers such as 0.6, 0.8, 0.9, 0.95, 0.99, 0.997, and so on, but regardless of how much time we spend going through those numbers and regardless of how many numbers we pass and examine, we will never reach the right end of the interval; we can never get to the last, largest number (the predecessor of 1) at the right end of the interval, because there simply isn't such a number. What's more, regardless of how close we are to the right end of this interval, there still is an uncountable infinity of real numbers between us and the end. The same is true of the left end of the interval. If we move toward it, we reach smaller and smaller numbers, we move steadily toward zero, but we can never reach the left end of the interval because that end (the smallest positive real number) does not exist.

Pick up two real numbers x and y at random (but with a uniform distribution) in the interval $(0, 1)$, divide them to obtain the real number $R = x/y$, and examine the integer I nearest R . We intuitively feel that I can be even or odd with the same

probability, but careful calculations [Weisstein-picking 07] show that the probability of I being even is 0.46460... instead of the expected 0.5.

A book may contain text, tables, mathematical expressions, and figures, but regardless of its contents it can be converted to PDF format and stored as a computer file. Such a file, like any data file, can be considered an integer or a long string B of digits (decimal, binary, or to any other base). A real number is also a (finite or infinite) string of digits. Thus, it is natural to ask, is there a real number that includes B in its string of digits? The answer is yes. Even more, there is a real number that includes in its exact, infinite representation all the books ever written and all those that will ever be written. Simply generate all the integers (we will use binary notation) 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, 0001, ... and concatenate them to construct a real number R . From its construction, R includes every possible bitstring and thus every past and future book. (Students pay attention. Both the questions and answers of your next examination are also included in this number. It's just a question of finding this important part of R .) However, R is a relatively small number because we can write it as

$$R = 0.01|00\ 01\ 10\ 11|000\ 001\ \dots\ 111|0000\ \dots,$$

and it is mind boggling to realize that this infinite amount of knowledge can be stored in a number that is less than 1.

The term Lexicon generally refers to a dictionary, but in mathematics, a lexicon is a real number that contains in its expansion infinitely many times anything imaginable and unimaginable, everything ever written, or that will ever be written, and any descriptions of every object, process, and phenomenon, real or imaginary. Contrary to any intuitive feelings that we may have, such monsters are not rare. The surprising result, due to [Calude and Zamfirescu 98], is that almost every real number is a Lexicon! This may be easier to comprehend by means of a thought experiment. If we put all the reals in a bag, and pick out one at random, it will almost certainly be a Lexicon.

Here is another baffling property of real numbers, the case of repeating 9's. The repeating real number $0.999\dots$ —which is also denoted by $0.\bar{9}$, $0.\dot{9}$, and $0.(9)$ —is especially interesting. At first look it seems to be less than 1. However, it is easy to show (and to prove rigorously in many ways) that this real number equals 1. Here are several simple proofs:

1. Denote $a = 0.\bar{9}$. The number a cannot be greater than 1, so it must be either less than 1 or equal to 1. Assume that $a < 1$, then there must be numbers between a and 1, such as, for example, $b = (a + 1)/2$. A little thinking shows that b cannot exist, because there is no way to write it. The number a already employs the digit 9, which is the largest decimal digit, so b cannot use any larger digits. Also, a contains infinitely many 9's, so b cannot have any more 9's in its representation. Thus, b cannot exist, which implies that a cannot be less than 1 and must therefore be 1.

2. Using fractions, it is easy to see that

$$1 = \frac{3 \times 1}{3} = 3 \times \frac{1}{3} = 3 \times 0.\bar{3} = 0.\bar{9}.$$

3. Manipulating digits is also a simple procedure that proves our claim. Denote $a = 0.\bar{9}$, we obtain $10a = 9.999\dots$, and $10a - a = 9.999\dots - 0.999\dots$ or $9a = 9 \Rightarrow a = 1$.

4. A geometric series has the form $ar + ar^2 + ar^3 + \dots$. It is known that if the absolute value of r is less than 1, the sum of the series is $ar/(1 - r)$. Our number $0.\bar{9}$ can be written as the geometric series

$$9 \left[\frac{1}{10} \right] + 9 \left[\frac{1}{10} \right]^2 + 9 \left[\frac{1}{10} \right]^3 + \dots$$

where r is 0.1. Its sum is therefore the finite quantity

$$\frac{9 \left[\frac{1}{10} \right]}{1 - \frac{1}{10}} = 1.$$

These proofs demolish a long-held intuitive belief among students of mathematics, namely that a real number has one representation. It is now clear that infinitely many numbers have multiple representations. Thus, 52.8357 can also be written as the infinite-repeating number $52.8356\bar{9}$ and $1/3$ also equals $0.\bar{3}$, $2/6$, $4/12$, and so on.

The remainder of this discussion is devoted to the concept of computable and uncomputable real numbers, because it is this concept that leads to doubts about the existence (or the reality) of most reals.

Imagine the real interval $[0, 1]$. This is a short segment where each point is associated with a real number. It seems a simple concept, but it becomes strange (even mysterious) when we try to compute such a number digit by digit. Many years of research, thoughts, and discussions have resulted in the conclusion that most real numbers cannot be computed digit by digit with a discrete, finite algorithms; such numbers are uncomputable and therefore remain nameless and elusive; mere ghosts.

In 1936–37, Alan Turing published his classical paper [Turing 37] “On computable numbers, with an application to the Entscheidungsproblem.” In this document, which is currently considered the foundation of theoretical computer science, Turing introduces a discrete model of computation which is now known as the Turing machine. This model is then used by Turing to discuss the concept of computability by considering computable and uncomputable numbers (i.e., real numbers that can be computed by a discrete, finite algorithm running on a Turing machine or on any general-purpose computer, and real numbers that cannot be computed in principle). We are familiar with a few real numbers—such as π , e , and ϕ —and they can be computed digit by digit (which is also true of every other real number that we have seen).

In his paper, Turing points out that there are infinitely many possible algorithms for computing real numbers, but this infinity is countable, which is why an algorithm can be tagged or numbered (first algorithm, second one, etc.). The number of possible algorithms is countable because an algorithm is realized by a computer program and any program can be considered an integer (the concatenation of the ASCII codes of its individual characters). Thus, even if there are infinitely many computer programs, their number is countable. The reals, however, are uncountable, so there must be (infinitely) many real numbers that are uncomputable.

Entscheidungsproblem is a German term meaning “Decision Problem.” This term refers to the problem of deciding whether a given assertion is a theorem or not, within the bounds of an axiomatic system. This problem originated first by Leibniz in the 17th century and then by David Hilbert in 1928 as part of his plan of constructing a formal axiomatic system that encompasses all of mathematics. Kurt Gödel, and a few years later, also Alan Turing, showed that for any mathematical system capable of encompassing the whole of arithmetic, there will exist an unsolvable Entscheidungsproblem, i.e., an assertion can be constructed whose truth or falsity cannot be decided within the system; in other words, undecidable statements exist for any axiomatic system.

In his paper, Turing even constructs an uncomputable real number. He employs the familiar diagonal method originated by Georg Cantor. Arrange all the algorithms for computing real numbers in an infinite (but countable) vertical list. (Even simpler, include every computer program, not just those that compute real numbers, in this list.) Run all these algorithms. Delete from the list any algorithms that do not generate numbers. The result is an infinite vertical list of real numbers. Take the first (leftmost) digit of the first number R_1 , the second digit of the second number R_2 , and so on. Change each digit in some way and concatenate the results to form a real number R . It is easy to see that R is none of the numbers on our vertical list because its first digit is not the first digit of R_1 , its second digit is not the second digit of R_2 , and so on. Therefore, R must be uncomputable.

This is an elegant way to construct an uncomputable number, but people have long ago pointed out that R is actually computable because we have shown how to determine each of its digits.

Consider the one-unit interval $[0, 1]$. Geometrically, it is a short, straight segment, one of the simplest geometric figures. It does not seem cause any conceptual difficulties or to contain any mysteries or any unusual numbers. However, Emile Borel showed that if we pick a number at random from this interval, the probability that it will be computable is infinitesimal (although not zero). Borel (perhaps together with Henri Lebesgue) also showed that any infinite countable set of real numbers does not contain any computable reals. Thus, in order to have any chance of finding a computable real, we have to have an infinite uncountable set of reals (which is easy, since any interval has uncountably many reals).

In 1927, Borel published a short, four-page paper describing another uncomputable number (this paper appears as an appendix to several late editions of his book, see [Borel 50]). All possible French texts, even those that are infinitely long, make up a countable set. Therefore all possible questions with a yes/no answer that can be asked in French constitute a countable set. Thus, the questions can be tagged (or numbered) and their answers, each of which is a single bit, can be concatenated to form a single real number K . We can refer to K as a know-it-all number, and it has answers to all the possible yes/no questions, even to questions whose answers are still unknown (such as: is the Riemann hypothesis true? is there intelligent life elsewhere in the universe?). In a sense, this single number has an infinite amount of information, except that it cannot be constructed. Thus, to Borel, K was unreal; it did not exist.

Opponents of Borel found weak points in his argument. For example, there are yes/no questions, such as “is the answer to this question NO?” for which there is no answer. Thus, attempts have been made to extend Borel’s idea and make it more rigorous. One such attempt is the real number known as Ω , the brainchild of Gregory Chaitin, the originator of algorithmic information theory. In his book [Chaitin 07] and lecture [Chaitin 08], Chaitin describes a real number that is well defined and is a specific number, but is impossible to compute.

Computer programs are notoriously difficult to write and to fully debug. So much so that many software experts claim that a large program can never be completely debugged. One common behavior of an undebugged program is infinite loops. A program with such a loop will never terminate. (The behavior of a program depends on its input data, so we assume self-contained programs, programs without input; any data needed by the program is contained as constants within the program.) Clearly, it is useful to know beforehand whether a given program will ever stop, and this question became known as the halting problem. Given a program, find whether or not it is going to halt, without actually running it. (It actually does not help to run the program, because it may halt after many years and we simply cannot wait that long.) In [Turing 37] it is proved that this problem is undecidable. There is no discrete, finite algorithm that will decide whether a given program will ever terminate.

In fact, a program may run forever even if it is fully debugged. Here is a simple example. All prime numbers, except 2, are odd. When two odd integers are added, the result is even. Thus, the sum of any two primes is even. In the 18th century, Christian Goldbach examined the opposite of this statement. Is it true that every even integer is the sum of two primes? Goldbach tried many even integers, and others after him used computers to check many more. So far, the answer has always been yes, so this question (whose answer is still unknown) has become known as the Goldbach conjecture. A computer program that checks this conjecture will either run forever checking more and more even integers, or will halt when it finds the first even number that is not the sum of two primes.

Now imagine the countable list of all computer programs. Some programs on this list will terminate, while others will run forever. This associates a single bit with each program, 0 for a non-terminator and 1 for a program that eventually halts. The set of bits is considered a real number that Chaitin named Ω . This is a number whose n th digit (in base 2) tells whether the n th computer program stops or not.

Such a number exists, because any computer program halts or does not halt, but the clever feature that distinguishes Ω from Turing’s R is that Ω , while fully specified and defined, cannot be constructed digit by digit, because it is impossible in principle to tell whether a given program will terminate. Thus, Ω is uncomputable.

It turns out that the definition of Ω can be improved, because in the original definition, the bits of Ω are not completely independent; they feature some redundancy. Imagine a finite set S of computer programs (i.e., a set of bits in Ω). If we know in advance that n of these programs halt, then we can run all the programs in S and wait until n of them stop. We will then know that the remaining programs will never terminate. Thus, the set of bits that correspond to S has redundancy. If S contains N programs, then of the N bits it contributes to Ω , only $\log_2 N$ are independent.

Those bits in Ω that correspond to S are correlated and can therefore be compressed. Once compression is achieved, the resulting Ω is maximally uncomputable (and also maximally unknowable), because its bits are compressed and therefore seem random, in spite of their being defined in a precise and simple way. It is as if this number contains an infinite amount of information, a notion that may disturb certain people and also seems incompatible with a finite universe.

Instead of using an existing compression algorithm, we change the meaning of Ω in a way that yields a maximally-compressed number, i.e., a number without redundancy. We define the new Ω as the probability that a program chosen at random will halt. Ω is now defined as an infinite sum, to which each N -bit program that halts contributes the term $1/2^N$. (Each N -bit program that halts adds a 1 to the N th bit in the binary expansion of Ω . The final value is the infinite sum of all the $1/2^N$ contributions from the programs that halt. The length of a program depends on the programming language, so for each language there is a different Ω , but all these numbers are uncomputable and have the same unusual properties.

Here is a simple example. We look at a computer program as a bitstring, whose precise value depends on the language. We construct all the possible programs (i.e., bitstrings) by selecting the next bit of a program at random and then running the program. Suppose that only the following programs halt, 101, 11011 and 11101. Their lengths are, 3, 5 and 5 bits. There are eight 3-bit programs, so the chance of generating any of them at random is $1/8 = 1/2^3$, and similarly for the 5-bit programs. Thus, the chance of generating any of the three halting programs by this random process (or, equivalently, the chance that the computer will halt when running any program) is the sum $1/2^3 + 1/2^5 + 1/2^5 = 0.001 + 0.00001 + 0.00001 = 0.0011_2 = 0.1875_{10}$ and this number is Ω for our example.

A technical note. In defining Ω we consider only self-delimiting programs, programs that stop asking for more bits when they halt. Once we know that program 101 halts, we ignore any longer programs that start with 101 and we do not add any terms to the sum for those programs. We regard all the longer programs as being included in the halting of 101.

Chaitin shows that Ω is fully compressed by showing that one cannot use a program substantially shorter than N bits long to compute the first N bits of Ω .

The resulting Ω is not only uncomputable, but is maximally so, because its bits look random, they feature no structure or pattern and can be said to be irreducible. This single number contains infinite irreducible complexity and it illustrates the problem we face when trying to understand the reals.

Ω embodies an enormous amount of wisdom in a very small space... inasmuch as its first few thousands digits, which could be written on a small piece of paper, contain the answers to more mathematical questions than could be written down in the entire universe.

—From [Bennett and Gardner 79]

This discussion causes some people to question the reality of real numbers with infinite precision. It seems that such numbers have certain questionable, problematical aspects, aspects that become more serious the deeper we study real numbers, their

existence, and how to compute them.

Now, if we take the modern approach to science, an approach that says that something is real if it can be observed or computed, then the unavoidable conclusion is that the uncomputable reals are not real (pun intended). Chaitin always asks, if a real number cannot be computed, then in what sense does it exist? We can say that the majority of the reals are unreal. They are mathematical shadows, phantoms that cannot be named, cannot be referred to, and cannot be isolated. They are ghosts that we have created in our imagination in order to fill up empty positions in the real number line; in any case, not real objects.

Reference [Plouffe 09] is a server of 215 million mathematical constants, many of which are real numbers.

Unusual, unexpected, counterintuitive, infinite, also unreal? The real numbers are weird, which is why some consider them beautiful, intriguing, and well worth the effort to gain a deeper understanding!

The following joke (from [Renteln and Dundes 05]) has long become a permanent part of mathematical lore:

Q: How many mathematicians does it take to screw in a lightbulb?

A: 0.999999...

References

Bennett C. H. and M. Gardner (1979) “The random number omega bids fair to hold the mysteries of the universe,” *Scientific American*, **241**:20–34.

Borel, Emile (1952) *Leçons sur la Théorie des Fonctions*, Gabay, Paris, pp. 161, 275.

Calude, Cristian and Tudor Zamfirescu (1998) “The Typical Number is a Lexicon,” *New Zealand Journal of Mathematics*, **27**:7–13.

Chaitin, Gregory (2007) *Meta Maths: the Quest for Omega*, Atlantic Books.

Chaitin, Gregory (2008) “How real are the real numbers?” a talk given at the 2008 Midwest NKS conference. This is file `confplanning.html` at <http://www.cs.indiana.edu/~dgerman/2008midwestNKSconference/>

Renteln, Paul and Allan Dundes (2005) “Foolproof: A Sampling of Mathematical Folk Humor,” *Notices of the AMS*, **52**(1):24–34. Available at <http://www.ams.org/notices/200501/fea-dundes.pdf>.

Plouffe (2009) is <http://pi.lacim.uqam.ca/eng/>.

Turing, A. (1937) “On computable numbers, With an Application to the Entscheidungsproblem,” *Proceedings of the London Mathematical Society*, **42**(2):230–265. This is (commercially) available in PDF at <http://plms.oxfordjournals.org/content/vols2-42/issue1/index.dtl>.

Weisstein-pickin (2007) is <http://mathworld.wolfram.com/RealNumberPicking.html>
Weisstein, Eric W. “Real Number Picking.” From MathWorld, A Wolfram web resource.