WHAT IS A REAL NUMBER?

JOHN MYHILL, State University of New York at Buffalo

In this paper I shall try by examples to give some of the feeling of constructive mathematics. I shall adopt the point of view of Bishop, which is in many ways clearer than that of Brouwer, the originator of constructive mathematics. I shall consider the notion of real numbers from a constructive point of view. This point of view requires that any real number can be *calculated*. It does not believe in the existence of any object which has not been constructed. We shall explain various senses in which it can be said that a real number has been constructed, and explain why some of these are unsuitable for the purpose of developing analysis constructively.

As a first approximation, let us say that a real number has been constructed if a rule has been given which enables us to compute its nth decimal place for any positive integer n. The notion of a "rule" is a primitive one in constructive mathematics, but it must be understood that the application of a rule is a mechanical matter; no intelligence is involved. In particular we may think of a digital computer, which given any positive integer n, will print out the number $f(n)$, as defining the rule f. In fact nobody has ever given an example of a function from positive integers to positive integers which can be calculated in a mechanical way, other than those which can be calculated by suitably idealized digital computers—the so-called *recursive functions*. Thus in practice it might suffice to identify rule-like functions of natural numbers with recursive functions. This identification, however, does not in our opinion belong to mathematics but to philosophy, and we shall abstain from making it. We therefore take the notion of a rule as an undefined one; in practice we seem to be able always to recognize when a mechanical process has been described.

From the constructive point of view, the only functions which exist are those which have been constructed; that is, functions for whose evaluation a rule has been given. For example, if we define

$$f(x) = \begin{cases} 0, \text{ if } a^x + b^x \neq c^x \text{ for all} \\ \text{integers } a, b, c > 0, \\ 1, \text{ if } a^x + b^x = c^x \text{ for some} \\ a, b, c > 0, \end{cases}$$

John Myhill received his Harvard Ph.D. under W. V. Quine. He has since held positions in philosophy and mathematics departments at Vassar, Temple, Yale, Chicago, Berkely, Stanford, Illinois, and presently SUNY at Buffalo. He held a Guggenheim Fellowship while at Chicago, and he spent three fellowship years at the Institute for Advanced Study. He spent the last academic year as a visitor at the Univ. of Leeds, England. He is the author of numerous papers on recursive functions, foundations of mathematics, and computer science; also of *Recursive Equivalence Types* (with J. C. E. Dekker, Univ. of Calif. Press, 1960) and of *Intuitionism and Proof Theory* (co-editor with Kino and Vesley, North Holland, 1969). *Editor.*

we have defined, from the classical point of view, a function. However, from a constructive point of view this does not constitute a definition of a function, because no directions have been given for computing it.

Because of the restriction to rule-like functions, we shall henceforth use the words 'rule' and 'function' interchangeably. Thus we shall not regard f just defined as being a function at all.

Our first attempt at explaining what is meant by a real number is then as follows: α is a real number if a rule has been given to compute the nth decimal place of α. Thus a real number α can be identified with a function ϕ from non-negative integers to integers, where $\phi(0)$ is the integer part of α and where for $n > 0$, $\phi(n) \in \{0, \cdots, 9\}$. We shall denote the set of real numbers in the sense of this definition by R_d (d for "decimal").

Although most of the real numbers encountered in analysis (for example all the algebraic numbers, and the transcendental numbers e and π) are constructible in this sense, we shall show that the set R_d is not suitable as a foundation for analysis. In fact we prove the following disagreeable thing:

THEOREM 1. *The set R_d of real numbers possessing a decimal expansion is not closed under addition, i.e., there are numbers α, $\beta \in R_d$ such that the number $\alpha + \beta$ is not in R_d.*

Before I prove this, I must explain the constructive sense of the word "not". This is used in *historical* sense; that is, to say that a proposition is not true means that no one has yet proved it. From the constructive point of view, just as nothing exists until it has been constructed, so no proposition is true until it has been proved. Constructivists reject the idea that in some platonic realm a T or an F has been placed beside each mathematical proposition P, independently of whether anyone knows whether P is true. There is another constructive notion resembling "not", called *absurdity*: P is called *absurd* if the assumption of P yields a contradiction. The notion of absurdity shares some of the properties of the classical "not", but it does not, for example, satisfy the law of excluded middle; it is simply untrue that for every proposition P, P has either been proved or shown to be absurd (contradictory). The law of excluded middle, "P or not P" in the classical sense, appears to the constructivist to be a piece of mythology; it says that in some non-material world, truth-values have already been assigned to all propositions, independent of human mathematical activity. Constructivists cannot make sense of this third kind of "not"; a truth that nobody knows how to prove makes as little sense to the constructivist as a real number that nobody knows how to calculate.

In Theorem 1, "not" is used in the historical sense. We shall give two numbers α and β such that each of them can be computed to any required number of decimal places, while yet nobody knows even the first decimal place of $\alpha + \beta$. To prove this (historical) assertion, we shall use our (historical) ignorance of the behavior of the decimal expansion of π. Specifically, nobody knows whether a sequence 5555 occurs in that expansion. If such a sequence occurs beginning at the kth place, and if it is the first such sequence, then k is called the *critical number* (of π). Nobody knows whether such a number exists, and nobody knows whether (if it exists) it is even or odd. Further, given any nonnegative integer n, one can evidently determine whether n is critical or not; all one has to do is compute the first $n + 3$ decimal places of π.

Now I give directions for computing the decimal expansions of the numbers α and β.

To compute α we write down

$$.33333\ldots\ldots\ldots$$

and continue writing 3 unless we reach some *odd* place, $2n + 1$, such that $2n + 1$ is the critical number of π. In that case we write a 4 at the $2n + 1$-st place and ever afterwards.

Thus if the critical number k of π is odd, $\alpha > \frac{1}{3}$, but if k is even or does not exist, $\alpha = \frac{1}{3}$.

To compute β we write down

$$.66666\ldots\ldots\ldots$$

and continue writing 6 unless we reach some *even* place, $2n$, such that $2n$ is the critical number of π. In that case we write a 5 at the $2n$th place and ever afterwards.

Thus if the critical number k of π is even, $\beta < \frac{2}{3}$, but if k is odd or does not exist, $\beta = \frac{2}{3}$; ("k does not exist" means "no 5555 occurs in the decimal expansion of π").

We have

$$\text{if } k \text{ is even } \alpha = \tfrac{1}{3},\ \beta < \tfrac{2}{3},\ \alpha + \beta < 1;$$

$$\text{if } k \text{ is odd } \alpha > \tfrac{1}{3},\ \beta = \tfrac{2}{3},\ \alpha + \beta > 1; \text{ and}$$

$$\text{if } k \text{ does not exist, } \alpha = \tfrac{1}{3},\ \beta = \tfrac{2}{3},\ \alpha + \beta = 1.$$

Now suppose we could write down even one place of the decimal expansion of $\alpha + \beta$. Then

$$\text{if } \alpha + \beta \text{ begins } 1 \cdot \cdots, \text{ then } \alpha + \beta \geqq 1$$

and if k exists, it is odd; while

$$\text{if } \alpha + \beta \text{ begins } \cdot 9 \cdots, \text{ then } \alpha + \beta \leqq 1$$

and if k exists, it is even.

Thus if we could compute even one place of $\alpha + \beta$, we could prove one of the two propositions "if k exists, it is odd" or "if k exists, it is even." That is, we could either prove "if 5555 occurs in π, its first occurence begins at an odd place," or else we could prove "if 5555 occurs in π, its first occurence begins at an even place." But we have not proved either of these two propositions; thus we cannot write down even one decimal place of $\alpha + \beta$, even though we can write down all the places of α and β. This completes the proof of Theorem 1.

Now we consider another possible approach to real numbers. One can object to the above proof that it is artificial; it uses the numbers α and β that are not *located* with respect to the rationals. To say a real number λ is located with respect to the rationals is to say that we can decide, for every rational number r, which of the three alternatives $\lambda < r$, $\lambda = r$, $\lambda > r$, holds. Thus α is not located with respect to $\frac{1}{3}$, β is not located with respect to $\frac{2}{3}$, and $\alpha + \beta$ is not located with respect to 1. We shall also require that we know an integer upper bound M on $|\lambda|$. This enables us to compute the decimal expansion of any located real number λ. For we first compare λ with each of the integers

$$-M, -M+1, \cdots, 0, \cdots, M-1, M$$

to get the whole number part of λ, say q; then we compare λ with each of $q + \frac{1}{10}$, $q + \frac{2}{10}, \cdots, q + \frac{9}{10}$, to get the first place after the decimal point, and so on. The situation is as follows:

THEOREM 2. *Let R_l (l for "located") denote the set of all located real numbers. Then $R_l \subset R_d$, but the converse does not hold.*

$R_l \subset R_d$ we have just proved. To disprove $R_d \subset R_l$, we must give a number with a decimal expansion which is not located with respect to the rationals. The number α of the preceding theorem is such a number. For we showed how to compute its successive decimal places, but we have not proved any of the three propositions "$\alpha < \frac{1}{3}$," "$\alpha = \frac{1}{3}$," or "$\alpha > \frac{1}{3}$." ($\alpha < \frac{1}{3}$ is absurd since every digit of α is either 3 or 4; $\alpha = \frac{1}{3}$ would imply "if k exists it is even," and $\alpha > \frac{1}{3}$ would imply "k exists and is odd." But we have not proved either of these propositions.) Hence $\alpha \in R_d - R_l$.

The condition of being located is therefore strictly stronger than that of having a decimal expansion. Furthermore, most of the real numbers encountered in analysis are located—the algebraic numbers for example, and the numbers e and π, as was shown by Goodstein. By way of illustration the number $\sqrt{2}$ is located; for to determine whether a rational r is $<$ or $>$ $\sqrt{2}$ ($=$ of course is impossible), we simply ask first if $r \leqq 0$; if it is, then $r < \sqrt{2}$, if not we compute r^2 and ask if $r^2 <$ or > 2. In fact we can prove a stronger property of $\sqrt{2}$ which we shall need in the sequel.

THEOREM 3. *For any rational number r, we can compute a number n_r such that* $|r - \sqrt{2}| > 1/10^{n_r}$. (This means that the decimal expansion of r differs from that of $\sqrt{2}$ at or before the n_rth place.)

Proof. We have

$$|r - \sqrt{2}| \geqq \big||r| - \sqrt{2}\big| = \frac{|r^2 - 2|}{|r| + \sqrt{2}} > \frac{|r^2 - 2|}{|r| + 2}.$$

So pick n_r so large that $1/10^{n_r} < |r^2 - 2|/(|r| + 2)$.

It will probably be felt that any reasonable number is located, and that the fact that R_d is not closed under addition results from the fact that numbers like α and β in Theorem 1 are not located. This might incline us to define computable real numbers λ as located rather than decimally expandible real numbers; formally, as pairs (N,f), where $N > |\lambda|$ and where for each rational r, $f(r) = 0$, 1, or 2, according as $\lambda <, =,$ or $> r$. But this too will not do since we have more trouble:

THEOREM 4. *R_l is not closed under addition, i.e., there exist numbers γ, $\delta \in R_l$ such that $\gamma + \delta \notin R_l$.*

Proof. Let $\gamma \equiv \sqrt{2}$. We do not know whether 5555 occurs in the decimal expansion of $\sqrt{2}$. Define 'critical number of $\sqrt{2}$' as we defined 'critical number of π' before. To compute δ, write down the decimal expansion of $\sqrt{2}$, except that if n is the critical number of $\sqrt{2}$, we write 0 at the nth place and thereafter. Clearly $\gamma \in R_l$. Clearly also $\gamma + \delta \notin R_l$. For if $\gamma + \delta = 0$, $\delta = \sqrt{2}$ and no 5555 occurs in the decimal expansion of $\sqrt{2}$; while if $\gamma + \delta < 0$, such a 5555 does occur. Thus if $\gamma + \delta$ were located with respect to 0 we could determine whether $\sqrt{2}$ possesses a critical number, which we cannot. It remains to prove $\delta \in R_l$.

Let then a rational number r be given; we must show how to decide $r < \delta$, $r = \delta$, or $r > \delta$. First find if $r > \sqrt{2}$ or $r < \sqrt{2}$.

CASE I. $r > \sqrt{2}$. Then certainly $r > \delta$, for $\delta \leqq \sqrt{2}$.

CASE II. $r < \sqrt{2}$. By Theorem 3 we can find n such that r and $\sqrt{2}$ differ at or before the nth decimal place. Let n_0 be the least such n. The n_0th place of r is less than n_0th place of $\sqrt{2}$. Then if (SUBCASE II.1) there is no critical number of $\sqrt{2} \leqq n_0$, $\sqrt{2}$ and δ agree for their first n_0 places and $r < \delta$. If on the other hand (SUBCASE II.2) there is a critical number $\leqq n_0$, we can compute β exactly and compare it with r directly. In any case we can decide whether $r <, =,$ or $> \delta$ and so $\delta \in R_l$.

So we cannot use R_l as a foundation for analysis. We now give another definition which avoids the above difficulties. A *finite decimal* is a number of the form $a/10^b$, where a is an integer and b is a nonnegative integer; a real number ρ is called *decimally approximable* ($\rho \in R_{da}$) if given any rational $\varepsilon > 0$ we can find a finite decimal d with $|\rho - d| < \varepsilon$. This is wider than either of the preceding notions.

THEOREM 5. *$R_d \subset R_{da}$, but the converse implication does not hold.*

Proof. Let $\rho \in R_d$. Then by definition we can compute any desired number of places of the decimal expansion of ρ. To approximate it within $1/10^n$ we need only compute $n + 1$ places; hence $\rho \in R_{da}$. To refute the converse observe that the sum of two elements of R_{da} is again in R_{da}. For if d_1 and d_2 are decimal $\varepsilon/2$-approximations to ρ_1 and $\rho_2 \in R_{da}$, then

$$|(d_1 + d_2) - (\rho_1 + \rho_2)| \leqq |d_1 - \rho_1| + |d_2 - \rho_2| < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

so that $d_1 + d_2$ is a decimal ε-approximation to $\rho_1 + \rho_2$. Hence $\rho_1 + \rho_2 \in R_{da}$. Now let α, β be as in Theorem 1, $\alpha, \beta \in R_d$ but $\alpha + \beta \notin R_d$. Then $\alpha, \beta \in R_{da}$ and so $\alpha + \beta \in R_{da} - R_d$.

Thus the decimally approximable real numbers form a more likely candidate as a foundation for constructive analysis than either R_d or R_l. The following theorem confirms this impression:

THEOREM 6. *R_{da} is a field.*

We have just proved that R_{da} is closed under addition; as an example of the verification of the remaining field postulates we shall prove that it is closed under multiplication. Let $\rho_1, \rho_2 \in R_{da}$: we seek a decimal ε-approximation to $\rho_1\rho_2$. We first compute (from 1-approximations to ρ_1 and ρ_2) a number $M > \max(|\rho_1|, |\rho_2|)$. Now find $\varepsilon/2M$-approximations d_1, d_2 to ρ_1, ρ_2 respectively, with $|d_1|, |d_2| < M$. We have

$$|d_1 - \rho_1|, |d_2 - \rho_2| < \varepsilon/2M$$

$$\begin{aligned} |d_1d_2 - \rho_1\rho_2| &= |d_1(d_2 - \rho_2) + \rho_2(d_1 - \rho_1)| \\ &< M|d_2 - \rho_2| + M|d_1 - \rho_1| \\ &< M(\varepsilon/2M) + M(\varepsilon/2M) = \varepsilon, \end{aligned}$$

so that d_1d_2 is an ε-approximation to $\rho_1\rho_2$.

In verifying the field postulates, we have to make sure that the statement of some of them makes constructive sense. For example, in the postulate

$$(*) \qquad x \neq 0 \rightarrow (\exists y)(xy = 1)$$

we must be careful to give the right meaning to the hypothesis $x \neq 0$. It is easy to construct a number which is neither $<, =,$ or > 0. For example, the number $\gamma + \delta$ in Theorem 4 is such a number (recall that if $\gamma + \delta = 0$, no 5555 occurs in the decimal expansion of $\sqrt{2}$; if $\gamma + \delta < 0$, such a 5555 does occur, while $\gamma + \delta > 0$ is absurd). Now $x \equiv \gamma + \delta \in R_{da}$, but x is neither $<, =,$ or > 0. How are we to construe (*) for such an x? The correct version is: If x is *separated* from zero, i.e., if a rational number r with $0 < r < |x|$ is known, then x possesses a reciprocal. This notion of separation is an example of how constructive mathematics (except in counterexamples) normally replaces negative statements by positive ones.

It may come as a surprise to some to learn that R_{da} is a *complete* field, in the sense that if $\{\rho_i\}$ is a sequence of elements of R_{da} such that for every $\varepsilon > 0$ we can compute N_ε with

$$|\rho_i - \rho_j| < \varepsilon \qquad (i,j > N_\varepsilon),$$

then we can construct a number $\lim \rho \in R_{da}$ satisfying

$$(\forall\varepsilon)(\exists M_\varepsilon)(\forall i > M_\varepsilon)|\rho_i - \lim \rho| < \varepsilon.$$

The proof is in fact a rather straightforward computation with ε's and δ's.

Of course R_{da} is not an *ordered* field; we just saw an example of an element $x \equiv \gamma + \delta$ of R_{da} which was neither $>, <,$ or $= 0$. However, R_{da} is closed with respect to all the usual functions connected in analysis, and indeed is sufficiently like the classical continuum that Bishop has made it the foundation of his book on constructive analysis. What is more remarkable is that the arguments of his book (not the counterexamples, but the theorems) are to an unexpected extent scarcely different from the classical ones. When they differ, they surpass the classical ones in precision and numerical content; for example, the proofs of existence always contain a method for approximating the number asserted to exist.

I conclude with two remarks of a more specialized nature. Firstly, I would like to make precise the difference between *constructive* analysis and *recursive* analysis. What we have been doing is constructive analysis; it admits no real numbers other than computable ones and no methods of proof other than constructive ones, and the notion of "computable function" or "rule" is a primitive one. Recursive analysis (e.g., in the sense of Klaua) on the other hand, is the study, by whatever means one wishes, of a certain classically defined subset of the real numbers, called the recursive reals. "Computable" is simply a synonym for "recursive" and is a defined idea. From the point of view of what I call recursive analysis, the sets R_d, R_l and R_{da} are all the same, but the proof that they are the same is non-constructive.

My last remark concerns the *formalization* of the remarks in this paper. If one takes a two-sorted theory, with variables for natural numbers and computable functions, and postulates, for the former, Peano's axioms and (primitive) recursive definition and for the latter, simply the axiom of choice

$$(\forall x)(\exists y)A(x,y) \rightarrow (\exists f)(\forall x)A(x,f(x));$$

and if the underlying logic is taken to be the intuitionistic predicate calculus, I think one has an adequate foundation for the constructive theory of real numbers. (Of course that is not the whole of constructive analysis; for the theory of functions of a real or complex variable one needs functionals of higher types for which one also postulates axioms of choice and the possibility of primitive recursive definition. But for our purposes it is enough to consider just the simple two-sorted theory mentioned.) Note that the notion of "recursive" or "computable" function does not appear at all; the function-variables range *only* over computable functions.

How, finally, is one to formalize in this theory the counter-examples we have been discussing? One possibility is to adjoin *rules of rejection* as well as rules of proof; for example let $P(x)$ denote 'x is the critical number of π', then we postulate $Px \bigwedge Py \rightarrow x = y$,

$$P(x) \bigvee \neg P(x)$$

(P is decidable) and assert that both of the formulas $(\forall x)(Px \rightarrow x \text{ is even})$ and $(\forall x)(Px \rightarrow x \text{ is odd})$ are to be rejected (as not yet proved). The rules of rejection are: if $A \vdash B$ and B is rejected, then A is rejected: if A and B are rejected, so is $A \bigvee B$; if $A(x)$ is rejected (for all x) so is $(\exists x)A(x)$. On this basis we can formally prove that the two inclusions $R_d \subset R_l$ and $R_{da} \subset R_d$ are rejected.