Aspects of Constructivism

by

Errett Bishop

Notes on the lectures
delivered at the Tenth Holiday Mathematics Symposium
held at New Mexico State University, Las Cruces,
during the period December 27-31, 1972

These lectures will be different from most mathematical talks, because much of what I am going to say is disputatious. I could avoid this aspect of the subject, or anyway minimize it, by confining my remarks to the technical aspects of constructivism. This is precisely what I do not want to do, and not from any desire to promote controversy, either. Since I shall be presenting my own version of constructivism, although there are many others, I hope that some of you will take issue with me as we go along. Some of you may even wish to defend classical mathematics! It is time to bring the fundamental issues of mathematics, that have been hidden from public view so long, out into the open. These issues are not complex, and suited to study by experts only, as the experts would have us believe. As an instance of a simple but fundamental issue, how can there be numbers that are not computable by any known method? Does that not contradict the very essence of the concept of number, which is concerned with computation?

Such questions, striking at our lifelong conception of mathematics, may make these lectures seem excessively negativistic. It all depends on what you think of negativism, and how you use it.

To take an example, most of you have heard of non-standard analysis. A number of people who at various times have approached me in order to express an interest in constructivism have gone on to spoil it by expressing an interest in non-standard analysis as well. The two are at opposite poles. Constructivism is an attempt to deepen the meaning of mathematics; non-standard analysis, an attempt to dilute it further.

Mathematicians do not like to be told that their theorems are deficient in meaning. However there is the consolation that mathematics stands well in a relative sense. Anyone who has been around a university for a few years must have observed that some disciplines regard any non-superficial concern with meaning as bad form. The object is to do research that conforms to the rules of the game, any meaningful results being incidental. The question before us here is, to what extend has mathematics fallen into this trap? The unavoidable consideration of this question gives constructive mathematics its negativistic cast. The critique of classical mathematics (meaning mathematics as currently practiced by almost all pure mathematicians) that I shall present here was given its present incisive form by Brouwer. Once Brouwer's critique is out of the way, we shall be able to turn our attention to positive developments, which by now constitute 99% of the subject. In introductory lectures such as these, the philosophical foundations must be laid, and critical comments come to the fore.

The process that has diluted the meaning of mathematics can be observed in a broader context. The context can be as broad as life itself. Whether in mathematics, in economics, in sociology, or in social intercourse, to give a few instances, the attainment of meaningful expression is extremely difficult. It seems to me that our problems arise from our attempts to insulate ourselves from these difficulties. This insulation is achieved by a codification of correct procedure, by a methodology. When methodology is elevated to dogma, as it always is to some extent, attention is diverted from meaning and shifted to more formal ground.

Mathematics itself affords a striking instance of the transition from methodology to dogma. The axiomatic method, having been dormant for centuries, has become extremely active in recent times. It accounts for much that is powerful and profound, and I have no quarrel with its use in mathematics proper. I do take exception to its use in mathematical philosophy. The experts routinely equate the entire panorama of mathematics with the productions of this or that formal system. Philosophising about mathematics consists of manipulating formal systems! Although the lack of adequate attention to meaning in classical mathematics preceded the rise of formalism, dogmatic philosophising has greatly impeded the development of constructivism. These remarks do not mean that formal systems have no value. I am saying that they have been promoted as something they are not, as powerful tools for investigating the nature of mathematics and even as the font of meaning.

Since I have led into the problem of meaning, let us be more precise about that term. Mathematics has meaning on at least four levels:

(a) as a game-an intellectual challenge-like chess

(b) as an art form-a beautiful structure-like music

(c) as a tool-for understanding and manipulating nature

(d) as a description-of certain abstract entities.

Some prominent mathematicians are quick to say that for them it is just a game. A much more natural and compelling game than chess, needless to say. Let us call such people formalists. The typical formalist is an elder statesman, who no longer tries to make sense of it all. It is disturbing that there are more and more precocious formalists, who embrace the mathematics-as-a-game philosophy relatively early in their careers.

Of course, it is a game. It is also an art form. However, many of us think that its essence lies deeper. Here I should distinguish pure mathematics from applied, because their essences are distinct. Hermann Weyl once defined mathematics as "a branch of the theoretical construction of the one real world." This is an elegant and delightful definition, if one is speaking of applied mathematics. To get a definition of pure mathematics, I would turn Weyl's definition around, and define pure mathematics to be "that component of our precise intellectual activity which in principle is independent of the one real world." Be that as it may, the non-formalists will agree that the essence of applied mathematics derives from its utility, and the essence of pure mathematics from its descriptive content. For the present, let us put applied mathematics aside, and philosophize about pure mathematics. If we accept that its essence derives from its descriptive content, then what is being described? Alas, there is disagreement. The classicist wishes to describe God's mathematics; the constructivist, to describe the mathematics of finite beings, man's mathematics for short. (You notice I am not being quite fair. I don't see why I should be fair. Nobody else is.) The classical point of view is also called Platonism, or idealism, and the constructive, realism. To make the distinction clear, let us consider three theorems.

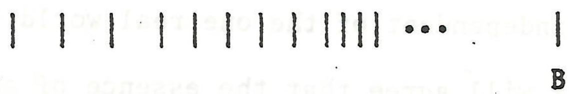
I. Every positive integer is the sum of 4 squares.

II. Every bounded monotone sequence of real numbers converges.

III. There exist irrational numbers x and y such that x^y is rational.

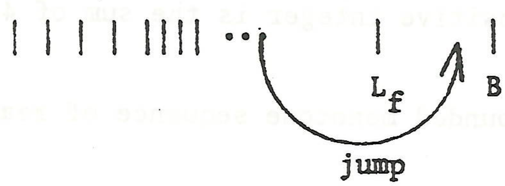
Theorem I is a beautiful example of man's mathematics. A finite being, given enough time, can represent an arbitrary positive integer as the sum of four squares.

Theorem II is a beautiful example of God's mathematics. There is no way a finite being can compute the limit of an arbitrary bounded monotone sequence. I will discuss this matter in more detail later. For the moment, a picture will make the point. The terms of the sequence are represented by vertical marks marching to the right, but remaining to the left of the bound B.



The classical intuition is that the sequence gets cramped, because there is only a finite amount of space left to it to the left of B, but infinitely many terms. Therefore the sequence has to pile up somewhere. That somewhere is its limit L.

I grant that this is precisely the behavior of some sequences. I call those sequences stupid. Let me tell you what a smart sequence will do. It will pretend that it is stupid, piling up at a limit L_f. Then when you have been convinced that that is actually what it is doing, it will jump to the right of L_f!



Of course, the smart sequence can only outwit you and me. It cannot outwit God. Thus Theorem II is false constructively but true classically.

Let me show you an amusing proof of Theorem III. Either $\sqrt{2}^{\sqrt{2}}$ is rational or it is irrational. If $\sqrt{2}^{\sqrt{2}}$ is rational, Theorem III is proved. If $\sqrt{2}^{\sqrt{2}}$ is irrational, take $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$. Then x and y are irrational and $x^y = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^2 = 2$ is rational!

Of course, God will be satisfied with this proof. He can look at $\sqrt{2}^{\sqrt{2}}$ and tell you right away whether it is rational or not. We are not able to do this. Therefore, unless we append to this proof a method by which a finite being can decide whether $\sqrt{2}^{\sqrt{2}}$ is rational, we do not have a constructive proof.

The fact that our proof of Theorem III was not constructive does not mean that Theorem III itself is not constructive. In fact, it is not hard to give Theorem III a constructive proof. Perhaps I shall do so later, after we have given a constructive proof of Cantor's result on the uncountability of the real numbers. For now, I only wish to make the point that it is only the given proof of Theorem III that is unconstructive, in contrast to Theorem II, which is essentially unconstructive, as the above considerations indicate. In classical mathematics, one considers the meaning of a theorem to be independent of its proof. As Theorem III indicates, to the constructivist the meaning of a theorem depends very much on its proof!

How do you know whether a proof is constructive? If you don't like the man or God point of view, look at it this way. Try to write a computer program. If you can program a computer to do it, it should be constructive. Notice I said write the program. Don't necessarily run it on the computer and wait around for the result.

The requirement of computability can be expressed more precisely as follows.

Fundamental Constructivist Thesis

Every (representation of) an integer can be converted in principle to decimal form by a finite, purely routine, process.

As the statement indicates, we work with representations of integers, not integers. What an integer is, I don't know, and I don't care. For simplicity, in the sequel I shall not bother to distinguish an integer from its representations.

Let us examine three classically defined (representations of) integers, and see how they measure up to our thesis. Let n_1 be 0 if every even integer between 4 and 10^4 is the sum of two primes, and 1 otherwise. Let n_2 be 0 if every even integer between 4 and 10^{100} is the sum of two primes, and 1 otherwise. Let n_3 be 0 if every even integer ≥ 4 is the sum of two primes, and 1 otherwise.

Although I have not personally looked into the matter, or know of anyone who has, it is not beyond my powers to compute n_1. Without even doing so, I am willing to bet the result will be 1.

In principle, the computation of n_2 (i.e., its conversion to decimal form) is equally simple. A novice could write the computer program. You see why I was careful to insert the phrase "in principle" into the constructivist thesis!

There is no known finite method for converting n_3 to decimal form. As some of you may have noticed, the representation of n_3 in decimal form is equivalent to the solution of Goldbach's famous conjecture: n_3 is 0 if the conjecture is true and 1 if it is false. Until we have a finite method that will lead to a proof or disproof of Goldbach's conjecture, we are (constructively speaking) not entitled to accept n_3 as an integer. If somebody finds such a method, and appends it to the description of n_3 given above, the resulting total description will be the definition of an integer.

There isn't really much more to tell you about integers. Two of them are equal, of course, if their decimal representations are equal in the usual sense. Their ordering and their arithmetic are also defined in terms of their decimal representations.

Let's move on to the rational numbers. The constructivist thesis is easily extended: Every (representation of) a rational number can be converted in principle to the form $\frac{p}{q}$, where p and q are decimal integers with $q \neq 0$, by a finite, purely routine, process. Equality, order, and arithmetic of rational numbers are defined in the usual way, working with their decimal representations $\frac{p}{q}$.

Let us move on, and ask what is meant constructively by a function $f:\mathbb{Z} \to \mathbb{Z}$ (where $\mathbb{Z}$ is the set of integers). We improve the classical treatment right away-instead of talking about ordered pairs, we talk about rules. Our definition simply takes a function $f:\mathbb{Z} \to \mathbb{Z}$ to be a rule that associates to each (constructively defined) integer n a (constructively defined) integer $f(n)$, equal values being associated to equal arguments. For a given argument n, the requirement that $f(n)$ be constructively defined means that its decimal form can be computed by a finite, purely routine process. That's all there is to it. Functions

$f:\mathbb{Z} \to Q$, $f:Q \to Q$, $f:\mathbb{Z}^+ \to Q$ are defined similarly. (Here Q is the set of rational numbers and $\mathbb{Z}^+$ the set of positive integers.) A function with domain $\mathbb{Z}^+$ is called a sequence, as usual.

Now that we know what a sequence of rational numbers is, it is easy to define a real number. A real number is a Cauchy sequence of rational numbers! Again, I have improved on the classical treatment, by not mentioning equivalence classes. I will never mention equivalence classes. To be sure we completely understand this definition, let us expand it a bit. Real numbers are not pre-existent entities, waiting to be discovered. They must be constructed. Thus it is better to describe how to construct a real number, than to say what it is. To construct a real number, one must

(a) construct a sequence $\{x_n\}$ of rational numbers

(b) construct a sequence $\{N_n\}$ of integers

(c) prove that for each positive integer n we have

$$|x_i - x_j| \le \frac{1}{n} \text{ whenever } i \ge N_n \text{ and } j \ge N_n.$$

Of course, the proof (c) must be constructive, as well as the constructions (a) and (b). This raises the question, of just what is a constructive proof, in particular a constructive proof of a statement such as (c). I am not prepared to say. My feeling is, any argument that I find completely convincing is an acceptable proof. Any argument that is not completely convincing, on the other hand, leaves something to be desired as a proof. The question often comes up, whether a constructivist should accept, for instance, a construction of a real number in which the sequences (a) and (b) are constructively defined but the proof (c) is classical and not constructive. My feeling is, I would want to see the proof before deciding whether to accept it. I do not believe that I would find a proof that relied on the principle of the excluded middle, for example, completely convincing, but perhaps one should keep an open mind. At the moment the question is academic anyway, because nonconstructive proofs of the type in question have not arisen.

Two real numbers $\{a_n\}$ and $\{b_n\}$ (the corresponding convergence parameters (b) and proofs (c) are assumed without explicit mention) are said to be equal if for each positive integer k there exists a positive integer N_k such that $|a_n-b_n| \leq \frac{1}{k}$ whenever $n \geq N_k$. It can be shown that this notion of equality is an equivalence relation. Addition and multiplication of real numbers are also defined, just like they are defined classically. The order relation, on the other hand, is more interesting. If $a = \{a_n\}$ and $b = \{b_n\}$ are real numbers, we define $a < b$ to mean that there exist positive integers M and N such that $a_n \leq b_n - \frac{1}{M}$ whenever $n \geq N$. Then it is easily shown that $a < b$ and $b < c$ imply $a < c$, that $a < b$ implies $a-c < b-c$, and so forth. Some care must be exercised in defining the relation $\leq$. We could define $a \leq b$ to mean that either $a < b$ or $a = b$. An alternate definition would be to define it to mean that $b < a$ is contradictory. We shall not use either of these, although our definition turns out to be equivalent to the latter.

Definition. $a \leq b$ means that for each positive integer M there exists a positive integer N such that $b_n \geq a_n - \frac{1}{M}$ whenever $n \geq N$.

To make the choice of this definition plausible, I shall construct a certain real number H, called the Royden number.

$$H = \sum_{n=1}^{\infty} \alpha_n 2^{-n}$$

where $\alpha_n = 0$ in case every even integer between 4 and n is the sum of two primes, and $\alpha_n = 1$ otherwise. (More precisely, H is given by the Cauchy sequence $\{a_n\}$, with $a_n = \sum_{k=1}^{n} \alpha_k 2^{-k}$, and the sequence $\{N_n\}$ of convergence parameters, where $N_n = n$.) Clearly we wish to have $H \geq 0$. It certainly is according to the definition we have chosen. (The real number 0 of course is the Cauchy sequence of rational numbers all of whose terms are 0.) On the other hand, we would not be entitled to assert that $H \geq 0$ if we had defined $H \geq 0$ to mean that either $H > 0$ or $H = 0$, because the assertion "$H > 0$ or $H = 0$" means that we have a finite, purely routine method for deciding which; in this case, a finite, purely routine method for proving or disproving Goldbach's conjecture!

Most of the usual theorems about $\leq$ and $<$ remain true constructively, with the exception of trichotomy. Not only does the usual form "$a < b$ or $a = b$ or $a > b$" fail, but such weaker forms as "$a < b$ or $a \geq b$", or even "$a \leq b$ or $a \geq b$" fail as well. For example, we are not entitled to assert "$0 < H$ or $0 = H$ or $0 > H$" for the Royden number H. If we consider the closely related number $H' = \sum_{n=1}^{\infty} \alpha_{2n}(-2)^{-n}$, we are not even entitled to assert that "$H' \geq 0$ or $H' \leq 0$".

Since trichotomy is so fundamental, we might expect constructive mathematics to be hopelessly enfeebled because of its failure. This situation is saved, because trichtomy does have a constructive version, which of course is considerably weaker than the classical.

Theorem. For arbitrary real numbers a, b, and c, with $a < b$, either $c > a$ or $c < b$.

Proof. Choose integers M and N_0 such that $a_n \le b_n - \frac{1}{M}$ whenever $n \ge N_0$. Choose integers N_a, N_b, and N_c such that $|a_n - a_m| \le (6M)^{-1}$ whenever $n,m \ge N_a$, $|b_n - b_m| \le (6M)^{-1}$ whenever $n,m \ge N_b$, $|c_n - c_m| \le (6M)^{-1}$ whenever $n,m \ge N_c$. Set $N = \max\{N_0, N_a, N_b, N_c\}$. Since a_N, b_N, and c_N are all rational numbers, either $c_N < \frac{1}{2}(a_N + b_N)$ or $c_N \ge \frac{1}{2}(a_N + b_N)$. Consider first the case $c_N \ge \frac{1}{2}(a_N + b_N)$. Since $a_N \le b_N - M^{-1}$, it follows that $a_N \le c_N - (2M)^{-1}$. For each $n \ge N$ we therefore have

$$a_n \le a_N + (6M)^{-1} \le c_N - (2M)^{-1} + (6M)^{-1}$$

$$\le c_n + (6M)^{-1} - (2M)^{-1} + (6M)^{-1} = c_n - (6M)^{-1}.$$

Therefore $a < c$. In the other case, $c_N < \frac{1}{2}(a_N + b_N)$, it follows similarly that $a > b$. This completes the proof of the theorem.

Do not be deceived by the use of the word "choose" in the above proof, which is simply a carry-over from classical usage. No choice is involved, because M and N_0, for instance, are fixed positive integers, defined explicitly by the proof of the inequality $a < b$. Of course we could decide to substitute other values for the original values of M and N_0, if we desired, so some choice is possible should we wish to exercise it. If we do not explicitly state what choice we wish to make, it will be assumed that the values of M and N_0 given by the proof of $a < b$ are chosen.

It is constructively correct to state that the real numbers are a complete ordered field, if in the definition of order we merely replace the usual version of trichotomy with the form just proved!

Another important property of the order relations is the following.

Theorem. Let a and b be real numbers. Then $a \geq b$ if and only if $a < b$ is contradictory.

Proof. Assume that $a \geq b$ and $a < b$. Then there exist positive integers M and N such that $a_n < b_n - \frac{1}{M}$ whenever $n \geq N$. Also, there exists a positive integer N_1 such that $a_n \geq b_n - \frac{1}{M}$ whenever $n \geq N_1$. If we take $n = \max\{N, N_1\}$ it follows that $a_n < b_n - \frac{1}{M}$ and $a_n \geq b_n - \frac{1}{M}$. This contradiction shows that if $a \geq b$, then $a < b$ is contradictory.

Assume conversely that $a < b$ is contradictory. We wish to prove that $a \geq b$. Let M be any positive integer. Choose an integer N_a such that $|a_n - a_m| \leq \frac{1}{3M}$ whenever $n \geq N_a$ and $m \geq N_a$. Choose N_b similarly. Write $N = \max\{N_a, N_b\}$. Consider any integer $m \geq N$. Assume that $a_m < b_m - \frac{1}{M}$. Then for each $n \geq N$ we have

$$a_n \leq a_m + \frac{1}{3M} < b_m - \frac{1}{M} + \frac{1}{3M} \leq b_n + \frac{1}{3M} - \frac{1}{M} + \frac{1}{3M} = b_n - \frac{1}{3M}.$$

It follows that $a < b$. Since $a < b$ is contradictory, $a_m < b_m - \frac{1}{M}$ is contradictory. Since this is an inequality between rational numbers, $a_m \geq b_m - \frac{1}{M}$. Since m is an arbitrary integer $\geq N$, it follows that $a \geq b$, as was to be proved.

Two real numbers x and y are said to be unequal, $x \neq y$, in case $x < y$ or $x > y$. This is easily seen to mean that there exist positive integers M and N with $|x_n - y_n| \geq \frac{1}{M}$ for all $n \geq N$.

Let r be any rational number. Then the sequence $\{r, r, r, \cdots\}$ is

Cauchy. Together with the sequence $\{N_k\}$ of convergence parameters, with $N_k = 1$ for all k, it defines a real number. This real number is also denoted by r. More generally, any real number that is equal to such a real number is also called rational. As with all constructive concepts, we would like irrationality to be a positive concept. We therefore define a real number x to be irrational if $x \neq r$ for all rational numbers r (rather than defining x to be irrational if it is contradictory that x is rational).

The Royden number H, which is constructively a well-defined real number, is classically rational, because if the Goldbach conjecture is true then $H = 0$, and if the conjecture is false then $H = 2^{-n+1}$, where n is the first even integer for which it fails. We are not entitled to assert constructively that H is rational: if it is rational, then either $H = 0$ or $H \neq 0$, meaning that either Goldbach's conjecture is true or else it is false; and we are not entitled to assert this constructively, until we have a method for deciding which. We are not entitled to assert H is irrational either, because if H is irrational, then $H \neq 0$, therefore Goldbach's conjecture is false, therefore H is the rational number 2^{-n+1}, a contradiction! Thus H cannot be asserted to be rational, although its irrationality is contradictory.

Later we shall prove the existence of many irrational numbers, by proving the uncountability of the real numbers, as a corollary of the Baire category theorem. For the present, let us merely remark that $\sqrt{2}$ is irrational. Of course, $\sqrt{2}$ can be defined by constructing successive decimal approximations. It is therefore constructively well-defined. The classical proof of the irrationality of $\sqrt{2}$ shows that if $\frac{p}{q}$ is any rational number then $\frac{p^2}{q^2} \neq 2$. Since both $\frac{p^2}{q^2}$ and 2 can be written with denominator q^2, it follows that

$$|\frac{p}{q} - \sqrt{2}| \cdot |\frac{p}{q} + \sqrt{2}| = |\frac{p^2}{q^2} - 2| \geq \frac{1}{q^2}.$$

Since clearly $\frac{p}{q} \neq \sqrt{2}$ in case $\frac{p}{q} < 0$ or $\frac{p}{q} > 2$, to show that $\frac{p}{q} \neq \sqrt{2}$ we may assume $0 \leq \frac{p}{q} \leq 2$. Then

$$|\frac{p}{q} - \sqrt{2}| \geq |\frac{p}{q} + \sqrt{2}|^{-1} \cdot \frac{1}{q^2} \geq |2+2|^{-1} \cdot \frac{1}{q^2} = \frac{1}{4q^2}.$$

Therefore $\sqrt{2} \neq \frac{p}{q}$. Thus $\sqrt{2}$ is (constructively) irrational.

The failure of the usual form of trichotomy means that we must be careful in defining absolute values and maxima and minima of real numbers. For example, if $x = \{x_n\}$ is a real number, with sequence $\{N_n\}$ of convergence parameters, then $|x|$ is defined to be the Cauchy sequence $\{|x_n|\}$ of rational numbers (with sequence $\{N_n\}$ of convergence parameters). Similarly, $\min\{x,y\}$ is defined to be the Cauchy sequence $\{\min\{x_n,y_n\}\}_{n=1}^{\infty}$, and $\max\{x,y\}$ to be $\{\max\{x_n,y_n\}\}_{n=1}^{\infty}$.

This definition of min, in particular, has an amusing consequence. Consider the equation

$$x^2 - xH' = 0.$$

Clearly 0 and the modified Royden number H' are solutions. Are they the only solutions? It depends on what we mean by "only". Clearly $\min\{0,H'\}$ is a solution, and we are unable to identify it with either 0 or H'. Thus it is a third solution! The reader might like to amuse himself looking for others. This discussion incidentally makes the point that if the product of two real numbers is 0 we are not entitled to conclude that one of them is 0. (For example, $x(x-H') = 0$ does not imply that $x = 0$ or $x - H' = 0$: set $x = \min\{0,H'\}$.)

A function from the set of real numbers $\mathbb{R}$ to $\mathbb{R}$ is simply a rule that associates to each real number x another real number $f(x)$, such that $f(x) = f(y)$ whenever $x = y$. Everything to be interpreted constructively, of course.

If a and b are real numbers with $a < b$, a function $f:[a,b] \to \mathbb{R}$ is defined similarly. It is called <u>continuous</u> if for each positive integer n there is a positive integer N_n such that $|f(x)-f(y)| \le n^{-1}$ whenever $|x-y| \le N_n^{-1}$. This concept is classically called uniform continuity, of course. It is equivalent classically to pointwise continuity. Constructively, however, there seems to be no way to deduce uniform continuity from pointwise continuity. Of course, uniform continuity is the basic concept. One might almost say that the only reason in classical mathematics for proving continuity of a function $f:[a,b] \to \mathbb{R}$ is to use the continuity to prove uniform continuity. Even this reason is specious, since all the usual proofs of continuity actually provide uniform continuity with very little additional work. Since the concept of pointwise continuity by itself is of little use, and since we are not able to use it to prove uniform continuity, the obvious thing to do is to define continuity to mean uniform continuity, as we have done above. Thus we have followed the spirit of the classical definition, rather than its letter, in constructivizing the classical notion of a continuous function $f:[a,b] \to \mathbb{R}$. Many people have told me that they thought it would have been better if in this and similar instances I had stuck to the letter of the classical terminology, even at the cost of doing violence to its spirit.

A continuous function $f:R \to \mathbb{R}$ is of course defined to be one whose restriction to every proper closed subinterval $[a,b]$ is (uniformly) continuous. By the word "proper" we mean $-\infty < a < b < +\infty$.

Working with the concepts already developed, we would have little trouble constructivizing the standard results of calculus. Rather than indicate the entire development, I shall consider one or two examples in detail, to get a feeling for the problems involved.

One classical theorem that fails, is that a continuous function on a closed interval $[a,b]$ attains its maximum. Let H' be the modified Royden number. For $0 \le x \le 1$ we set

$$f(x) = H'x.$$

Classically, f attains its maximum either at 0 or at 1, depending on whether $H' \le 0$ or $H' \ge 0$. Constructively, no soap. The most we can say is that l.u.b. f and g.l.b. f exist.

To be more precise about this latter statement, consider any set S of real numbers. By l.u.b. S, we mean a real number u such that (a) $x \le u$ for all x in S, and (b) for each positive integer n, there exists x_n in S with $x_n \ge u - \frac{1}{n}$. Notice the switch from the classical definition! We have preserved the spirit but violated the letter. If instead of (b) we had substituted the usual condition (b') $y \ge u$ for each upper bound y of S, then we would not have been able to prove (b) constructively. Since it is (b), and not (b'), that packs the punch, we require that (b) be satisfied. Going back to our function, l.u.b. f is defined as usual to be l.u.b.$\{f(x): x \in [a,b]\}$. We pass by the simple proof that l.u.b. f exists for all (uniformly!) continuous f.

Another theorem that fails constructively is the intermediate value theorem. To see this, let H' be the modified Royden number. For $-1 \le x \le 0$, we define $f(x) = x + (1+x)H'$. For $0 \le x \le 1$, we define $f(x) = H'$. For $1 \le x \le 2$, we define $f(x) = (2-x)H' + x-1$. We must be careful, because f is not defined on all of the interval $[-1,2]$. It is only defined on $D = [-1,0] \cup [0,1] \cup [1,2]$. It is not yet defined for the modified Royden number H', for instance. However, since f is uniformly continuous on D, the usual methods, which are constructively correct, provide a unique continuous extention to $[-1,2]$, which we continue to denote by f. Clearly $f(-1) = -1$ and $f(2) = 1$. By the intermediate value theorem, there exists x_0 in $[-1,2]$ with $f(x_0) = 0$. By a previous result, either $x_0 > 0$ or $x_0 < 1$. In the former case, $H' > 0$ is contradictory, and therefore $H' \le 0$. In the latter case, $H' < 0$ is contradictory, and therefore $H' \ge 0$. Thus from the intermediate value theorem we have deduced constructively that $H' \le 0$ or $H' \ge 0$. It follows that we are not entitled to assert the intermediate value theorem constructively.

The following weakened version is often useful.

<u>Theorem</u>. Let $f:[a,b] \to \mathbb{R}$ be continuous, with $f(a) \le f(b)$. In addition, assume that whenever a_0 and b_0 are real numbers with $a \le a_0 < b_0 \le b$, there exists a real number x with $a_0 \le x \le b_0$ and $f(x) \ne f(a_0)$. (This additional hypothesis can be paraphrased as "f is non-constant on each proper sub-interval of $[a,b]$".) Then for each real number c with $f(a) \le c \le f(b)$, there exists x in $[a,b]$ with $f(x) = c$.

<u>Proof</u>. We construct inductively a sequence $\{[a_n,b_n]\}$ of proper

sub-intervals of $[a,b]$, such that (a) $b_n - a_n \le (\frac{2}{3})^{n-1}(b-a)$, and (b) $f(a_n) \le c \le f(b_n)$. To begin, set $a_1 = a$ and $b_1 = b$. It will be enough to construct the next interval $[a_2,b_2]$, since the same process we use to construct $[a_2,b_2]$ from $[a_1,b_1]$ can be used to construct $[a_3,b_3]$ from $[a_2,b_2]$, and so forth. By assumption, there exists x with $\frac{a_1+b_1}{2} \le x \le \frac{1}{3}a_1 + \frac{2}{3}b_1$ such that $f(x) \ne f(\frac{a_1+b_1}{2})$. Therefore either $f(x) \ne c$ or $f(\frac{a_1+b_1}{2}) \ne c$. In either case, $f(y) \ne c$ for some y with $y - a_1 \le \frac{2}{3}(b_1-a_1)$ and $b_1 - y \le \frac{2}{3}(b_1-a_1)$. Since $f(y) \ne c$, either $f(y) < c$ or $f(y) > c$. In the former case, take $a_2 = y$ and $b_2 = b_1$. Then $b_2 - a_2 = b_1 - y \le \frac{2}{3}(b_1-a_1)$, and $f(a_2) = f(y) \le f(c) \le f(b_1) = f(b_2)$. In the latter case, take $a_2 = a_1$ and $b_2 = y$. Again, the required conditions are satisfied. This completes the inductive construction of $\{[a_n,b_n]\}$. Clearly $\{a_n\}$ is a Cauchy sequence. Since $\mathbb{R}$ is complete, it converges to a limit x. Clearly $\{b_n\}$ also converges to x. Since f is continuous and $f(a_n) \le c$ for all n, we have $f(x) \le c$. Similarly, $f(x) \ge c$. Therefore $f(x) = c$.

A useful application of this theorem is to a polynomial function. It can be shown that a non-constant polynomial function satisfies the additional hypothesis. Thus a polynomial function f with $f(a) < f(b)$ assumes all values c with $f(a) \le c \le f(b)$.

There is an entirely different sort of constructivization of the intermediate value theorem.

Theorem. Let $f:[a,b] \to \mathbb{R}$ be continuous, with $f(a) \le f(b)$. Then there exists a sequence $\{c_n\}$ of real numbers, such that for every real number

c with $c \neq c_n$ for all n, and $f(a) \leq c \leq f(b)$, there exists x in $[a,b]$ with $f(x) = c$. (In other words, all except countably many intermediate values are assumed.)

Proof. Let $\{d_n\}$ be an enumeration of the numbers $a + r(b-a)$, where r runs over the rational numbers in $[0,1]$. Consider any real number c with $c \neq c_n$ for all n, where $c_n = f(d_n)$. By induction, we construct a sequence $\{[a_n,b_n]\}$ of proper closed intervals of $[a,b]$, with $b_n - a_n = (\frac{1}{2})^{n-1}(b-a)$, and $f(a_n) \leq c \leq f(b_n)$. Take $a_1 = a$ and $b_1 = b$. Since $c \neq f(a_n)$ for all n, and $\frac{a+b}{2}$ is a member of the sequence $\{a_n\}$, we have $c \neq f(\frac{a+b}{2})$. Therefore $c > f(\frac{a+b}{2})$ or $c < f(\frac{a+b}{2})$. In the former case take $a_2 = \frac{a+b}{2}$ and $b_2 = b$. In the latter case take $a_2 = a$ and $b_2 = \frac{a+b}{2}$. It is clear that the required conditions are satisfied. A continuation of the same process gives an inductive construction of the sequence $\{[a_n,b_n]\}$. The rest of the proof is the same as the corresponding part of the proof of the previous theorem.

The two versions of the intermediate value theorem just given stand at two extremes. The first version has a stronger hypothesis than the classical result, but the conclusion is the same. It is therefore called an equal conclusion constructive substitute. The second version, on the other hand, is an equal hypothesis constructive substitute.

Before going further, we shall systematically re-examine the language of mathematics, to adapt it to the constructive point of view. The first to do this seems to have been Brouwer, whom we follow here. He remarked that the meanings customarily assigned to the terms "and", "or", "not", "implies", "there exists", and "for all" are not entirely appropriate to the constructive point of view, and he introduced more appropriate meanings where necessary.

The connective "and" causes no trouble. To prove "A and B", we must prove A and also prove B, as in classical mathematics. We have already discussed the connective "or". To prove "A or B" we must give a finite, purely routine method which after a finite number of steps either leads to a proof of A or to a proof of B. This is very different from the classical use of "or"; for example the statement "$H' \geq 0$ or $H' \leq 0$" is true classically, but we are not entitled to assert it constructively. The classical meaning of "or" is too vague to be of constructive use.

The connective "implies" is defined classically by taking "A implies B" to mean "not A or B". This definition would not be of much value constructively. Brouwer therefore defined "A implies B" to mean that there exists an argument which shows how to convert an arbitrary proof of A into a proof of B. Isn't this a more natural and intuitive definition anyway? To take an example, it is clear that "{(A implies B) and (B implies C)} implies (A implies C)" is always true constructively; the argument that converts arbitrary proofs of "A implies B" and "B implies C" into a proof of "A implies C" is the following: given any proof of A, convert it into a proof of C by first converting it into a proof of B and then converting that proof into a proof of C.

We define "not A" to mean that A is contradictory. By this we mean that it is inconceivable that a proof of A will ever be given. For example, "not $0 = 1$" is a true statement. The statement "$0 = 1$" means that when the numbers "0" and "1" are expressed in decimal form, a mechanical comparison of the usual sort checks that they are the same. Since they are already in decimal form, and the comparison in question shows they are not the same, it is impossible by correct methods to prove that they are the same. Any such proof would be defective, either technically or conceptually. As another example, "not (A and not A)" is always a true statement, because if we prove not A it is impossible to prove A--therefore, it is impossible to prove both.

Having changed the meaning of the connectives, we should not be surprised to find that certain classically accepted modes of inference are no longer correct. The most important of these is the principle of the excluded middle--"A or not A". Constructively, this principle would mean that we had a method which in finitely many, purely routine, steps would lead to a proof or disproof of an arbitrary mathematical assertion A. Of course we have no such method, and nobody has the least hope that we ever shall. It is the principle of the excluded middle--even a very special case thereof--that accounts for almost all of the important unconstructivities of classical mathematics. By this I mean that if one were to tack the principle of the excluded middle (or even a special case thereof, which I call the "limited principle of omniscience") onto the principles of inference that are constructively correct, he would have a system powerful enough to deduce almost all of the important results of classical mathematics. Another incorrect principle is "(not not A) implies A". In other words, a demonstration of the impossibility of the impossibility of a certain construction, for instance, does not constitute a method for carrying out that construction. This particular incorrect principle can of course be deduced from the excluded middle.

I could proceed to list a more or less complete set of constructively valid rules of inference involving the connectives just introduced. This would be superfluous. Now that their meanings have been established, the rest is common sense. As an exercise, show that the statement

$$\text{"}(A \to 0 = 1) \leftrightarrow \text{not } A\text{"}$$

is constructively valid.

The classical concept of a set as a collection of objects from some pre-existent universe is clearly inappropriate constructively. Constructive mathematics does not postulate a pre-existent universe, with objects lying around waiting to be collected and grouped into sets, like shells on a beach. The entities of constructive mathematics are called into being by the constructing intelligence. From this point of view, the very question "what is a set" is suspect. Rather we should ask the question, "What must one do to construct a set?". When the question is posed this way, the answer is not hard to find.

Definition. To construct a set, one must specify what must be done to construct an arbitrary element of the set, and what must be done to prove two arbitrary elements of the set are equal. Equality so defined must be shown to be an equivalence relation.

Notice that the sets we have constructed so far--the set of integers, the set of rational numbers, the set of sequences of integers, the set of real numbers, etc., all conform to the prescription just given. In the case of the real numbers, we took pains to present the definition in exactly the prescribed form, as the reader will see if he checks back!

A number of people have quarreled with this definition of a set as being too vague. This is probably because they have been unduly influenced by formalism. Our definition is no more or less vague than the classical definition of a set, as a collection of mathematical objects. If they find that too vague, and want to define a set in the context of a formal axiomatic system, then they have been unduly influenced by formalism.

While we are on the subject, we might as well define a function $f:A \to B$. It is a rule which to each element x of A associates an element $f(x)$ of B, equal elements of B being associated to equal elements of A.

The notion of a subset A_0 of a set A is also of interest. To construct an element of A_0, one must first construct an element of A, and then prove that the element so constructed satisfies certain additional conditions, characteristic of the particular subset A_0. Two elements of A_0 are equal if they are equal as elements of A.

For example, to construct an element of the set $\mathbb{R}^+$ of positive real numbers, we must construct a real number x and then prove that there exist positive integers M and N with $x_n \geq \frac{1}{M}$ whenever $n \geq N$.

Contrary to classical usage, the scope of the equality relation never extends beyond a particular set. Thus it does not make sense to speak of elements of different sets as being equal, unless possibly those different sets are both subsets of the same set. This is because for the constructivist equality is a convention, whose scope is always a given set; all this is conceptually quite distinct from the classical concept of equality as identity. You see now why the constructivist is not forced to resort to the artifice of equivalence classes!

After this long digression, consider again the quantifiers. Let $A(x)$ be a mathematical assertion depending on a parameter x ranging over a set S. To prove "$\forall x A(x)$", we must give a method which to each element x of S associates a proof of $A(x)$. Thus the meaning of the universal quantifier "$\forall$" is essentially the same as it is classically.

We expect the existential quantifier "$\exists$", on the other hand, to have a new meaning. It is not clear to the constructivist what the classicist means when he says "there exists". Moreover, the existential quantifier is just a glorified version of "or", and we know that a reinterpretation of this connective was necessary. Let the variable x range over the set S. Then to prove "$\exists x A(x)$" we must construct an element x_0 of S, according to the principles laid down in the definition of S, and then prove the statement "$A(x_0)$". We have already seen how this is sometimes impossible to do, even when we have already proved classically "$\exists x A(x)$", for example in the intermediate value theorem.

Again, certain classical uses of the quantifiers fail constructively. For example, it is not correct to say that "not $\forall x A(x)$ implies $\exists x$ not $A(x)$". On the other hand, the implication "not $\exists x A(x)$ implies $\forall x$ not $A(x)$" is constructively valid. I hope all this accords with your common sense, as it does with mine.

We are now in a position to give constructive versions of such abstract axiomatic systems as groups, topological spaces, manifolds, etc. We shall restrict ourselves to some simple results from the theory of metric spaces.

<u>Definition</u>. A metric space consists of a set E and a function

$\rho: E \times E \to \mathbb{R}$, such that

(a) $\rho(x,y) \geq 0$ for all x and y

(b) $\rho(x,y) = 0 \leftrightarrow x = y$

(c) $\rho(x,y) = \rho(y,x)$

(d) $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$.

The usual classical examples of metric spaces are also metric spaces in the constructive sense--the Euclidean spaces $\mathbb{R}^n$, the space $C([a,b])$ of continuous real-valued functions on a compact interval, and all subsets thereof, to give three examples.

The only problem in proving the Baire category theorem constructively is to state it correctly. Call a subset U of a metric space E open if for every x in U there exists $r > 0$ such that the open sphere $S(r,x)$ is a subset of U. Call a subset V of E dense if to each x in E and each $r > 0$ there exists a point y in V with $\rho(x,y) < r$. Call an intersection $\bigcap_{n=1}^{\infty} U_n$ of dense, open subsets of E a residual subset of E. (The term is not appropriate constructively, but we conform to classical usage.) The Baire category theorem goes as follows.

Theorem. A residual subset of a complete metric space E is dense in E.

Proof. Let $S(r_0,x_0)$ be any open sphere in E, and $\{U_n\}$ a sequence of dense open sets. We must construct a point x of $S(r_0,x_0) \cap \bigcap_{n=1}^{\infty} U_n$. To do this, we construct a sequence $\{S(r_n,x_n)\}$ of open spheres, such that

(a) $S(r_n,x_n) \subset S(\frac{1}{2} r_{n-1},x_{n-1}) \cap U_n$

(b) $r_n \leq 2^{-n}$

for all $n \geq 1$. This is done inductively as follows. Since U_n is dense, there exists a point $x_n \in S(\frac{1}{2} r_{n-1}, x_{n-1}) \cap U_n$. Since the intersection is open, there exists $r_n > 0$ such that (a) is satisfied. Replacing r_n by $\min\{r_n, 2^{-n}\}$, we see that (a) remains satisfied and (b) is satisfied as well. This completes the inductive construction. By (a) and (b),

$$\rho(x_n, x_{n-1}) \leq \frac{1}{2} r_{n-1} \leq 2^{-n}.$$

Therefore $\{x_n\}$ is a Cauchy sequence. Since E is complete, it converges to a limit x. By (a), $x_n \in S(\frac{1}{2} r_m, x_m)$ whenever $n > m$. In other words, $\rho(x_n, x_m) < \frac{1}{2} r_m$. Therefore $\rho(x, x_m) \leq \frac{1}{2} r_m$. Taking $m = 0$, we see that $x \in S(r_0, x_0)$. Taking $m \geq 1$ we see that $x \in S(r_m, x_m) \subset U_m$. Thus $x \in S(r_0, x_0) \cap \bigcap_{n=1}^{\infty} U_n$. This completes the proof.

Notice that a particular point of $S(r_0, x_0) \cap \bigcap_{n=1}^{\infty} U_n$ is constructed by the above proof. This is because at each stage of the induction the proof of the density of U_n selects a particular point x_n, and the proof of the openness of $S(\frac{1}{2} r_{n-1}, x_{n-1}) \cap U_n$ selects a particular value of r_n. This is a universal attribute of constructive existence proofs: such a proof always constructs a specific element.

We now apply the category theorem to show the uncountability of the real numbers $\mathbb{R}$. A set S is countable if there exists a sequence $\{s_n\}$ of elements of S, such that for each element s of S there exists a positive integer n with $s = s_n$. The usual proofs show that the set of integers $\mathbb{Z}$, the set of rational numbers Q, the products $\mathbb{Z} \times \mathbb{Z}$ and $Q \times Q$, for example, are countable. We wish to show that $\mathbb{R}$ is uncountable, but in a positive, useful sense, which turns out to be the following.

<u>Definition</u>. Let S be a set with an inequality relation $\neq$. Assume that to each sequence $\{s_n\}$ of elements of S there exists an element s of S with $s \neq s_n$ for each n. Then S is called *uncountable*.

Using the Baire category theorem, we can prove the following strengthened version of the uncountability of $\mathbb{R}$.

<u>Theorem</u>. Let $\{x_n\}$ be a sequence of real numbers. Then the set

$$U = \{x \in \mathbb{R} : x \neq x_n \text{ for all } n\}$$

is residual.

<u>Proof</u>. We need only show that each $U_n = \{x : x \neq x_n\}$ is dense and open. To show U_n is open, consider x in U_n. Set $r = |x-x_n|$. Then $S(r,x) \subset U_n$. Thus U_n is open. To show U_n is dense, consider x in $\mathbb{R}$ and $r > 0$. Then $x_n > x$ or $x_n < x + \frac{1}{2}r$. In the first case, define $y = x$. Then $y \in U_n$ and $|x-y| < r$. In the second case, define $y = x + \frac{1}{2}r$. Then $y \in U_n$ and $|x-y| < r$. Thus U_n is dense.

As a consequence, the irrational numbers are residual. This fact can be used to give a constructive proof of the existence of irrational numbers x and y such that x^y is rational (a classical proof was given earlier). We assume some facts from the constructive theory of the logarithmic and exponential functions. The function $f(u) = (\log_2 u)^{-1}$ takes the interval $(0,1)$ onto the interval $(-\infty,0)$. The inverse function $g(v) = 2^{1/v}$ takes $(-\infty,0)$ onto $(0,1)$. The equality $x^y = 2$ means that $y \log_2 x = 1$, or that $y = f(x)$. By the above theorem and the

Baire category theorem, there exists x in $(0,1)$ such that $x \neq r_n$ and $x \neq g(t_n)$ for all n, where $\{r_n\}$ is an enumeration of the rational numbers and $\{t_n\}$ an enumeration of the non-zero rational numbers. Thus both x and $y = f(x)$ are irrational. Also, $x^y = 2$.

Another pretty application of category theory is to construct transcendental numbers--in fact, to prove they are residual. One way to do this would be to show the algebraic numbers are countable, from which it follows as above that the transcendental numbers--those unequal to every algebraic number--are residual. Since the proof that the algebraic numbers are countable is a bit messy, we take another point of view, defining a real number x to be transcendental if $p(x) \neq 0$ for every polynomial p with integral coefficients not all of which are zero. Let $\{p_n\}$ be a listing of all such polynomials. It is easily seen that $\{x : p_n(x) \neq 0\}$ is open and dense for each n. The set of transcendental numbers, which is the intersection of all these sets, is therefore residual.

The two most important concepts in the theory of metric spaces are perhaps continuity and compactness. Each presents some problems to the constructivist, as Brouwer pointed out.

Since the problems presented by compactness are more easily resolved, let's start there. The usual definition in terms of open coverings having finite subcoverings won't do, because even the closed intervals don't seem to be provably compact. The same objection holds for sequential compactness. It is true, however, that the closed intervals are complete and totally bounded. Following Brouwer, we therefore define compact to mean complete and totally bounded. (To refresh your memories, a metric space X is totally bounded if for each $\varepsilon > 0$ there exists a finite sequence

$x_1,\cdots,x_n$ of points of X such that for each x in X one of the numbers $\rho(x,x_i)$ is less than ε. Such a finite sequence is called an ε-approximation to X.)

A very useful and very trivial result of classical mathematics is that a closed subset of a compact space is compact. This is not true constructively (where by closed subset we mean a subset that contains all of its limit points). For example, the subset $S = \{x \in [0,1]: x = 0,$ or $x = 1$ and the Riemann Hypothesis is true$\}$ of $[0,1]$ can not be asserted to be compact, although it is easily seen to be closed.

As an example of a non-routine constructivization of a classical result, we shall develop some theorems which are useful in some circumstances for constructing compact subsets of a compact metric space. Of course, of the two conditions involved in compactness--completeness and total boundedness--the latter is much more critical, because a totally bounded space can always be completed to become compact. As an example, if $x_1,\cdots,x_n$ are finitely many points of a compact space E, then the closure of the set $\{x_1,\cdots,x_n\}$ is a compact subset of E because it is obviously totally bounded. The following result gives less trivial compact subsets.

<u>Theorem</u>. Let E be a compact metric space, and x_0 a point of E. Let r be a positive real number. Then there exists a compact set $S \subset E$ such that

$$S(r,x_0) \subset S \subset S(8r,x_0).$$

<u>Proof</u>. We construct by induction subsets $S_1,S_2,\cdots$ of E, each consisting of a finite sequence of points, such that

(a) for each x in $S(r,x_0)$ there exists y in S_n with $\rho(x,y) \le 2^{-n+1}r$

(b) for each x in S_{n+1} there exists y in S_n with $\rho(x,y) \le 2^{-n+3}r$.

We begin the construction by taking $S_1 = \{x_0\}$. Clearly (a) is satisfied. Assume next that $S_1,\cdots,S_n$ have been constructed satisfying the relevant instances of (a) and (b). We shall show how to construct S_{n+1}. Let $x_1,\cdots,x_N$ be a $2^{-n}r$-approximation to E. For each i, $1 \le i \le N$, we define

$$\rho(x_i,S_n) = \min\{\rho(x_i,y_1),\cdots,\rho(x_i,y_k)\},$$

where $y_1,\cdots,y_k$ are the points of S_n. We can separate the points $x_1,\cdots,x_N$ into two groups such that $\rho(x_i,S_n) < 2^{-n+3}r$ for all x_i in the first group and $\rho(x_i,S_n) > 2^{-n+2}r$ for all x_i in the second. Let S_{n+1} consist of all those x_i in the first group. Clearly (b) is satisfied by S_{n+1}. To check (a), consider x in $S(r,x_0)$. By the inductive hypothesis, there exists y in S_n with $\rho(x,y) \le 2^{-n+1}r$. We also have $\rho(x,x_i) \le 2^{-n}r$ for some i. Therefore $\rho(x_i,S_n) \le \rho(x_i,y) \le \rho(x,x_i) + \rho(x,y) < 2^{-n+2}r$. It follows that x_i must be in the first group. Thus x_i belongs to S_{n+1}, and

$$\rho(x,x_i) \le 2^{-n}r = 2^{-(n+1)+1}r.$$

Thus (b) is satisfied. Let S be the closure of $\bigcup_{n=1}^{\infty} S_n$. Clearly $S(r,x_0) \subset S$ by (a). Consider positive integers $n \le m$, and a point y in S_m. Then

$$\rho(y,S_n) \le 2^{-n+3}r + \cdots + 2^{-(m-1)+3}r < 2^{-n+4}r$$

by (b). It follows that S is totally bounded. Since S is closed, it is compact. Taking $n = 1$, we also see that

$$\rho(y,x_0) = \rho(y,S_1) < 8r.$$

Therefore $\rho(y,x_0) \le 8r$ for all y in S. In other words, $S \subset S(8r,x_0)$. This completes the proof.

Before deepening this last result, we need to introduce the concept of continuity. Let $f:E \to E_1$ be a function from a compact metric space E to a metric space E_1. For reasons discussed above, we define continuity of such a function f to mean uniform continuity. Our basic tool for constructing compact subsets is the following.

Theorem. Let $f:X \to \mathbb{R}$ be a continuous real-valued function on a compact metric space X. Then there exists a sequence $\{a_n\}$ of real numbers, such that if a is any real number with $a \ne a_n$ for all n, then the set $X_a \equiv \{x \in X : f(x) \le a\}$ is compact.

Proof. Since X_a is clearly closed, we must only show it is totally bounded. Consider any positive integer n. Let $x_1,\cdots,x_{N_n}$ be a $\frac{1}{16}n^{-1}$-approximation to X. By the previous theorem, for $1 \le j \le N_n$ there exists a compact set X_{jn} such that

$$S(\tfrac{1}{16}n^{-1},x_j) \subset X_{jn} \subset S(\tfrac{1}{2}n^{-1},x_j).$$

Thus the compact sets X_{jn}, for $1 \le j \le N_n$, cover X, and $\rho(x,y) < n^{-1}$ for all x and y in X_{jn}. Choose such sets for each positive integer n. For each j and n, with $1 \le j \le N_n$, we write $c_{jn} = \text{g.l.b.}\{f(x) : x \in X_{jn}\}$. Let a be any real number with $a \ne c_{jn}$ for all j and n. We shall show that X_a is totally bounded, thereby proving the theorem. Let n be any positive integer. For each j,

$1 \le j \le n$, either $c_{jn} < a$ or $c_{jn} > a$. For each j with $c_{jn} < a$, choose a point x_j in X_{jn} with $f(x_j) < a$. Then these points x_j are a n^{-1}-approximation to X_a. To see this, consider x in X_a. Then $x \in X_{jn}$ for some j. Hence $c_{jn} \le f(x) \le a$. It follows that an x_j was chosen for this value of j. Also, $\rho(x,x_j) < n^{-1}$ since both x and x_j belong to X_{jn}. Thus the x_j are a n^{-1}-approximation to X_a. Hence X_a is totally bounded.

A useful application is to the function f on X defined by

$$f(x) = \rho(x,x_0)$$

for some fixed x_0 in A. The sets X_a are then the closed balls about x_0. By the theorem, all except countably many are compact. Using this result, one could construct partitions of unity.

We shall close these lectures with some general philosophical remarks. It may be necessary to dispel the impression that constructive mathematics is only interested in those numbers that somehow have to do with Goldbach's conjecture. Exactly the opposite. The only interest of such numbers is to use them to show that certain theorems of classical mathematics are not constructive. They play no more part in the positive development of constructive mathematics than they do in classical mathematics. A more precise critique of classical mathematics would not employ them at all, as we shall now demonstrate. By the "limited principle of ommiscience", I mean the statement that every sequence of integers either vanishes identically or has a non-zero term. If this principle (LPO for short) were valid constructively, we would be able to constructivize most classical results with little trouble. In other words, LPO accounts for a very large proportion of the unconstructivities of classical mathematics. There seems to be almost no hope of ever proving LPO constructively. It would provide finite routines for either proving or disproving Goldbach's conjecture, the Riemann hypothesis, Fermat's last theorem, because each of these unsolved problems is equivalent to the identical vanishing of a certain constructively defined sequence of integers.

Thus if we can show constructively that "A implies LPO", for a certain mathematical statement A, we not only know that A has not yet been proved constructively, but also that A is extremely difficult, most probably impossible, to prove constructively. The statement A that "$x = 0$ or $x \neq 0$ for every real number x" is an example. To show that $A \rightarrow$ LPO, let $\{x_n\}$ be any sequence of integers. Define $y_n = 0$ if $x_n = 0$ and $y_n = 1$ if $x_n \neq 0$. Then $y = \sum_{n=1}^{\infty} 2^{-n} y_n$ is a real number. The statement A implies that $y = 0$, which means $y_n = 0$ for all n, or that $y \neq 0$, which means that $y_n \neq 0$ for some n. Thus $A \rightarrow$ LPO. Therefore this particular statement A will most probably never be proved constructively. Other statements that imply LPO are the statement that every bounded monotone sequence of real numbers converges, and the statement that every real number is either rational or irrational.

A slight modification of LPO, called LLPO or the "lesser limited principle of ommiscience", is also useful for pointing out unconstructivities. LLPO states that "for every sequence of integers either the first non-zero term, if one exists, is positive, or else the first non-zero term, if one exists, is negative". It is easy to show that LPO $\rightarrow$ LLPO, but the converse does not seem to be possible to prove. Nevertheless, LLPO is almost equally as implausible as LPO. Since the intermediate value theorem and the dichotomy "every real number is either ≥ 0 or ≤ 0" each implies LLPO, it is very improbable either of these statements will ever be proved (constructively).

Many people feel that constructive mathematics does not go far enough. According to them, it should concern itself with computations that can be carried out in fact, not merely in principle. There is no objection whatever that I can see to doing this. Since the basic constructivist philosophy is to bring out all possible shades of meaning, the constructivist philosophy demands that the distinction between computable in principle and computable in fact (or practice) be made wherever possible. Since there is no sharp division between computations that can be performed in practice and those that cannot, what we really are asking for is some measure of the complexity, or difficulty of a computation. If some such measure can be obtained systematically, we should certainly build it into our mathematics. Until then, we will have to continue to treat such questions by ad hoc methods.

The distinction between computable in principle and not necessarily computable at all, on the other hand, is susceptible to systematic treatment, but not within the framework of classical mathematics. To treat this very important matter--of what is computable in principle--systematically, we must set up a new framework. This is the constructive framework, developed above.

The classical mathematician might object, that the constructive framework is also deficient: whereas the classical viewpoint inhibits the development of constructive meaning, the constructive viewpoint inhibits the development of classical meaning. This is untrue. Almost all classical theorems use only the principle of the excluded middle (E.M.) and, less frequently, the axiom of choice (A.C.) in addition to constructively correct principles in their proofs. These classical theorems can therefore be realized from the constructive viewpoint as implications, of the form "(E.M. and A.C.) $\rightarrow$ A", where A is the classical result. In most cases, we even have "LPO $\rightarrow$ A". Moreover this constructive realization does not do violence to the meaning, in my opinion. Thus classical mathematics can be regarded as a branch of constructive mathematics, whereas to regard constructive mathematics as a branch of classical mathematics is not possible. If the classical mathematicians would take this point of view, and write their theorems constructively as implications, say of the form "E.M. $\rightarrow$ A", rather than refusing to recognize the distinction between "A" and "E.M. $\rightarrow$ A", a significant advance in meaning would be achieved. They might even come to believe, as many constructivists do, that although theorems of the form "E.M. $\rightarrow$ A" are interesting, the real task of mathematics is to prove results not having that surplus baggage on the left.