Why is $e^{i\pi} = -1$?

I first saw the equation $e^{i\pi} = -1$ in a popular book on mathematics many years ago, probably sometime when I was in high school. In the book it was written in the form

$$e^{i\pi} + 1 = 0$$

which allowed the author to say that the equation related five of the most important constants in mathematics.

This statement intrigued me. The constants 1 and 0 were certainly very familiar to me, and I felt more or less comfortable with π at least for things like computing the area of a circle. I had some acquaintance with e. I knew it was roughly 2.7 and that it was used as the base of the natural logarithms. However, I had often wondered at the adjective "natural" in this context because it was clear to me that logarithms to the base 10 were much more natural. It seemed to me that it was much easier to work with a base whose integer powers were nice values like $10^2 = 100$ rather than $e^2 =$ some messy irrational value. However, I was willing to take it on faith that e was an important number and that one day I would know its true significance.

The constant i, on the other hand, was quite a mystery to me. I knew that it stood for the square root of -1, and I understood that it couldn't be one of the numbers I was familiar with (which I learned later were called real numbers) because all their squares were nonnegative. But if i was not a real number, what was it? I could not really conceive of a number that did not belong to the real numbers. I understood the reluctance of mathematicians of earlier eras to deal with "numbers" like i. How did one think about such things? Where did they fit into what I knew?

To compound my uncertainty as to what the equation $e^{i\pi} = -1$ really meant, i and π appeared as exponents. In my mind c^n for c a real number and n a positive integer meant multiply c by itself n times. On rare occasion n might be a fraction rather than an integer, but this boiled down to taking square roots and the like, so it didn't trouble me much. But what did it mean to raise c to the power π? The best I could do with this was reason that since π was between 3 and 4, then c^π was probably between c^3 and c^4. Using a table of logarithms (preferably to the base 10) I could even compute c^π for given c, which impressed me to no end since I really didn't quite understand what was going on.

But what to make of i as an exponent? This was surely beyond my comprehension. Since i had no place in my universe of numbers, I could see no way to think about raising some number to the power i. Thus, the equation $e^{i\pi} = -1$ was really a complete mystery to me.

As my mathematical education progressed, the light slowly began to dawn for me: The expression c^x is in itself meaningless, just a couple of letters of the alphabet positioned on the page in a certain way. Before it is possible to talk about this expression, much less discuss its value, we must *define* the expression, i.e., we must state clearly in terms of what we already know what this new notation represents.

Thus, for c a real number and x a positive integer, we *define* c^x to mean $c \cdot c \cdots c$, where c appears x times, or, somewhat less precisely, c^x is c multiplied by itself x times. Now we have a clear statement of what c^x means in terms of things we already know (assuming we know about multiplication of real numbers, the associativity of multiplication, etc.).

We can now prove facts about our new notation, e.g., $c^x c^y = c^{x+y}$. This is rather exciting; in some sense we are replacing multiplication by addition.

As we continue to work with this new idea, we may encounter c^0. What are we to make of this? After some thought we realize that c^0 is *undefined.* Our definition is for an exponent which is a positive integer, and zero is not a positive integer. Thus the expression c^0 is meaningless. We can't talk about it or do anything with it.

Actually, we *can* do something with it: We can *define* it to mean something. In principle, we can define it to mean anything we want, but if we are not careful, our definition will not be very useful.

Taking a moment to think about what we are really up to, we see that we are trying to extend an existing definition to a wider arena. We have a definition for c^x when x is a positive integer, and we want to extend this definition to apply when x is a nonnegative integer. With a bit more thought we realize that if the new definition is to be useful, then we want the important theorems we have proved about the existing definition to be true for the new extended definition also.

In particular, we want $c^x c^y = c^{x+y}$ to remain true when either x or y is zero. Thus, we want $c^x c^0 = c^{x+0}$. But this means that $c^x c^0 = c^x$, so, if c is not zero, we can cancel c^x on both sides to get $c^0 = 1$. It certainly looks like we should consider defining $c^0 = 1$.

At first we may resist this tentative definition. If $c^2 = c \cdot c$ and $c^1 = c$, shouldn't $c^0 =$ no c's at all? But what does this mean? Zero, perhaps? But if we say that $c^0 = 0$, then $c^x c^0 = c^x \cdot 0 = 0$, so $c^x c^0 = c^{x+0}$ is no longer true. In fact, we find that most of our interesting theorems about exponents are no longer true.

But, if we use the definition $c^0 = 1$, then the interesting theorems remain true. In mathematics this is a very strong incentive to use the definition. If we step back a bit and extract ourselves from the morass of pondering what a product of no c's at all might mean, hopefully we'll recall that we have complete freedom in defining what c^0 means. So why not define it to mean what seems to be most fruitful and useful? If subsequent work with the definition reveals that it is not so useful, well, then we can change the definition. (In fact, we may be forced to change or refine the definition; for example, what does 0^0 mean? Is it 1?)

Now that we've extended the definition of c^x once, why not keep going? In short order we can extend the definition to x any integer, x rational, and x real. The step from x rational to x real is of a very different nature than the other steps. It involves such things as limits and sequences of an infinite number of terms. It is in the process of taking this step that we discover why e is important and why it really is natural for e to be the base for logarithms.

In the generalization of c^x to wider and wider domains of x, we see the same theme played out each step of the way: We want to define c^x in the wider domain so that our existing theorems in the narrower domain are still true in the new domain. This results in a paradox of sorts: In principle we have complete freedom to define things any way we want. But, if we want our theorems to remain true in the new situation, we often have no freedom whatsoever. Plugging the new domain into the old theorems reveals what the new definition must be. We either accept this and make the required definition, or we must give up the theorems that have served us so well in the narrower domain.

There are instances where we refuse to accept the definition forced upon us by the old theorems. Quaternions are an example. Hamilton defined a generalization of complex numbers in which he gave up commutativity of multiplication, thus foregoing theorems such as $ab = bc \Rightarrow a = c$ for $b \neq 0$. It turned out in this case that giving up some old theorems was worthwhile. Hamilton started something that eventually became modern algebra and modern vector analysis.

But in many cases the most fruitful thing to do is to accept the definition forced upon us. This allows us to continue building a structure upon what we have already built without tearing down existing components.

We will illustrate this by an example, an investigation of how to extend the function e^x for real x to e^z, where z is a complex number.

We want to find a function $f(z)$ which maps complex numbers to complex numbers, and which satisfies the following conditions:

A. $f(x) = e^x$ for all real x

B. $f(z + w) = f(z)f(w)$ for all complex z and w

C. $f'(z)$ exists, i.e., $f(z)$ is analytic

We certainly want condition A to hold because otherwise $f(z)$ is not an extension of e^x. Condition B is the complex analog of the most important property of e^x, so we want it to be true also. Since e^x is analytic for real x, and since the theory of functions of a complex variable deals almost exclusively with analytic functions, we want condition C to hold too.

We regard A, B, and C as a bare minimum. Any $f(z)$ which has only one or two of these properties will not do. On the other hand, we can hope that we might find an $f(z)$ which not only satisfies these conditions, but in addition has the following properties:

D. $f'(z) = f(z)$

E. $f(z)$ is defined for all complex z

F. $|f(z)| \neq 0$ for all z, i.e., $f(z) \neq 0$

Properties D and E are exact analogs of properties of e^x. Since the complex numbers are not ordered, property F is as close as we can get to $e^x > 0$ for all real x.

Now, how do we begin to find an $f(z)$ with properties A, B, and C? One traditional method is to ask a passing psychic. If we are lucky enough to encounter the right psychic, we might be told to try

$$f(z) = e^x(\cos y + i \sin y), \tag{1}$$

where $z = x + iy$. It is fairly easy to show that all six of our conditions are met by this function. Thus, it not only fulfills the essential conditions A, B, and C, but it also has the hoped-for properties D, E, and F. So, we have found an extension of e^x to the complex plane.

However, our method leaves something to be desired. Often when one needs them most, there are no psychics in the neighborhood. We would like another approach, something that will show us how conditions A, B, and C lead us inevitably to (1).

Following [Bak and Newman, p37], we begin with $z = x + iy$ and consider $f(z) = f(x + iy)$. By condition B, $f(x + iy) = f(x)f(iy)$. By condition A, $f(x) = e^x$, so we have $f(z) = e^x f(iy)$. We can write $f(iy) = A(y) + iB(y)$ for some functions $A(y)$ and $B(y)$ which map reals to reals, so we end up with

$$f(z) = e^x A(y) + ie^x B(y) \tag{2}$$

Letting $u(x, y) = e^x A(y)$ and $v(x, y) = e^x B(y)$, and taking partial derivatives, we see that

$$\begin{aligned} u_x &= e^x A(y) & \qquad v_x &= e^x B(y) \\ u_y &= e^x A'(y) & \qquad v_y &= e^x B'(y) \end{aligned}$$

Applying the Cauchy-Riemann equations (which are equivalent to our condition C):

$$u_x = v_y \Rightarrow e^x A(y) = e^x B'(y) \Rightarrow A(y) = B'(y) \tag{3}$$

$$u_y = -v_x \Rightarrow e^x A'(y) = -e^x B(y) \Rightarrow A'(y) = -B(y) \tag{4}$$

Differentiating the last equation in (4) we get $A''(y) = -B'(y)$, and using the last equation in (3) this becomes $A''(y) = -A(y)$. All solutions of this differential equation are given by

$$A(y) = a\cos y + b\sin y \tag{5}$$

for a and b any real numbers. Since $B(y) = -A'(y)$, we get

$$B(y) = -b\cos y + a\sin y \tag{6}$$

Letting $z = x$, a real number, we see from (2) that

$$f(x) = e^x A(0) + ie^x B(0)$$

But $f(x) = e^x$, so it must be that $A(0) = 1$ and $B(0) = 0$. Plugging $y = 0$ into (5) and (6), we get $a = 1$ and $b = 0$, so $A(y) = \cos y$ and $B(y) = \sin y$. Thus,

$$f(z) = e^x(\cos y + i\sin y).$$

So, without the aid of a passing psychic, we see that conditions A, B, and C force us to define $f(z)$ as in (1). Furthermore, the proof shows that this $f(z)$ is unique, i.e., the conditions are satisfied by one and only one function.

It turns out that the $f(z)$ we seek is fully determined by rather weaker conditions than A, B, and C. The following theorem [p112 of Silverman] is true:

> There is a unique function $f(z)$ [which turns out to be (1)] with these properties:
>
> A'. $f(z)$ is defined for all complex z, $f(x)$ is real for all real x, and $f(1) = e$.
>
> B'. $f(z + w) = f(z)f(w)$ for all complex w, z.
>
> C'. $f(z)$ is differentiable at $z = 0$.

This is a surprising result. The function need only be differentiable at one point, and it need match e^x at only one point. The proof is rather long (4 pages) but it uses very elementary facts, little more than the Cauchy-Riemann equations.

Actually, an even more surprising fact is true: Condition B is not really needed. Given that $f(x) = e^x$ for real x and $f(x)$ is analytic, then $f(z)$ is fully determined. This can be seen by defining

$$f(z) = 1 + z + z^2/2! + z^3/3! + \dots,$$

in analogy with the series for e^x, using the fact that since $f(z)$ and e^x agree on the real line, $f(z)$ is the unique analytic continuation of e^x to the entire complex plane.

We can take yet another step: Let x_n be a sequence of real numbers containing an infinite number of distinct points, and let L be a limit point of the sequence. Then there is exactly one function $f(z)$ analytic in the complex plane such that $f(x_n) = e^{x_n}$ and $f(L) = e^L$. This function is (1). This is quite amazing.

What is equally amazing is that, in a sense, from the moment we defined c^x for x a positive integer, we completely determined the sequence of generalizations to x an integer, x rational, x real, and x complex. At each step we had virtually no choice in defining c^x in the new domain if we wished to preserve the laws of exponents. This must be regarded as an expression of the deep underlying unity of the number system, if not of mathematics itself.

We conclude with

$$e^{i\pi} = e^{0+i\pi} = e^0(\cos\pi + i\sin\pi) = 1(-1 + i\cdot 0) = -1,$$

answering the question posed in the title of this essay.

Acknowledgements

I would like to thank Prof. Rick Luttmann of Sonoma State University for inspiring this essay.

References

Bak, J. and Newman, D.J., *Complex Analysis*, Springer-Verlag, New York (1982).

Flanigan, F.J., *Complex Variables*, Dover Publications, Inc., New York (1972).

Knopp, K., *Theory of Functions, Part I*, Dover Publications, Inc., New York (1945).

Markushevich, A.I., *Theory of Functions of a Complex Variable*, three volumes in one, Chelsia Publishing Co., New York (1985) (available from the American Mathematical Society).

Silverman, R.A., *Introductory Complex Analysis*, Dover Publications, Inc., New York (1972).

Roger F. House
12655 Fiori Lane
Sebastopol, CA 95472
707-874-3258
rhouse@sonic.net
2007 Dec 28