

July 22, 2001

Mr. Charles I. Clausing
4 Cedar Falls Drive
Medford, NJ 08055

Dear Mr. Clausing,

Your letter of July 7 to Dean Jenkins, containing a proposed ruler-and-compass trisection of an angle, was referred to the Chair of the Mathematics Department, who then referred it to me. I was a bit busy last week and didn't get time to look at it until today. Let me give you the bottom line first; then you can decide if your cup is half empty or half full. Your construction is the simplest and most accurate *approximate* trisection I have ever seen using only ruler and compass. I expect "simplest and most accurate" may help to fill up your cup, and "approximate" may drain it a bit. Let me explain the issue further, before I analyze your construction.

The problem of trisecting an angle can be understood in more than one sense. There is the *practical* problem of giving a construction that is good enough for artistic and design purposes; and there is the *theoretical* problem of exhibiting a sequence of circles and lines whose intersections determine an angle equal to *exactly* one-third of a given angle (with *zero* error). Your trisection is of the first kind. The second kind does not apply to *physical* angles at all. The perfect circles, infinitely thin lines, points of zero length, and infinite precision involved are all products of human thought and "exist" (if the word is even appropriate in this context) only in the human mind.

I don't know how much background you have in this problem. Very often people who come to me with a trisection have heard only that the trisection is alleged to be impossible. I should emphasize that only the second (zero-error) trisection has been proved impossible. Very good approximate trisections such as yours have always been known to exist; no one has ever said they didn't. There are also zero-error trisections that can be made by bending the rules for using the ruler and compass a bit; one of those was discovered by Archimedes over two thousand years ago. But it really is impossible to do a zero-error trisection without bending the rules. This is explained in more detail in an essay I'm enclosing with this response, so you can read it if you like. In case you'd rather not, let me illustrate it by analogy. I would never believe anyone who told me the Miami Dolphins had defeated the Pittsburgh Steelers by a score of 2 to 1. That's not because I think the Steelers are a better team, but because I know it is impossible for the losing team in an American football game to score exactly 1 point. The rules simply don't allow it. If it really did happen, it would mean the rules had been changed. That's what's going on with the impossibility proof.

That being said, let me emphasize again that your approximate trisection is the simplest I've seen and astonishingly accurate. Now let's have a look at your argument. Here's a summary of what you said, referring to Fig. 1.

1. Start with any angle ABC
2. Bisect angle ABC with BD
3. Bisect angle ABD with BE
4. Bisect angle DBC with BF
5. Draw equal chords AD , EF and DC
6. Draw GE and HF parallel to BD extended to M

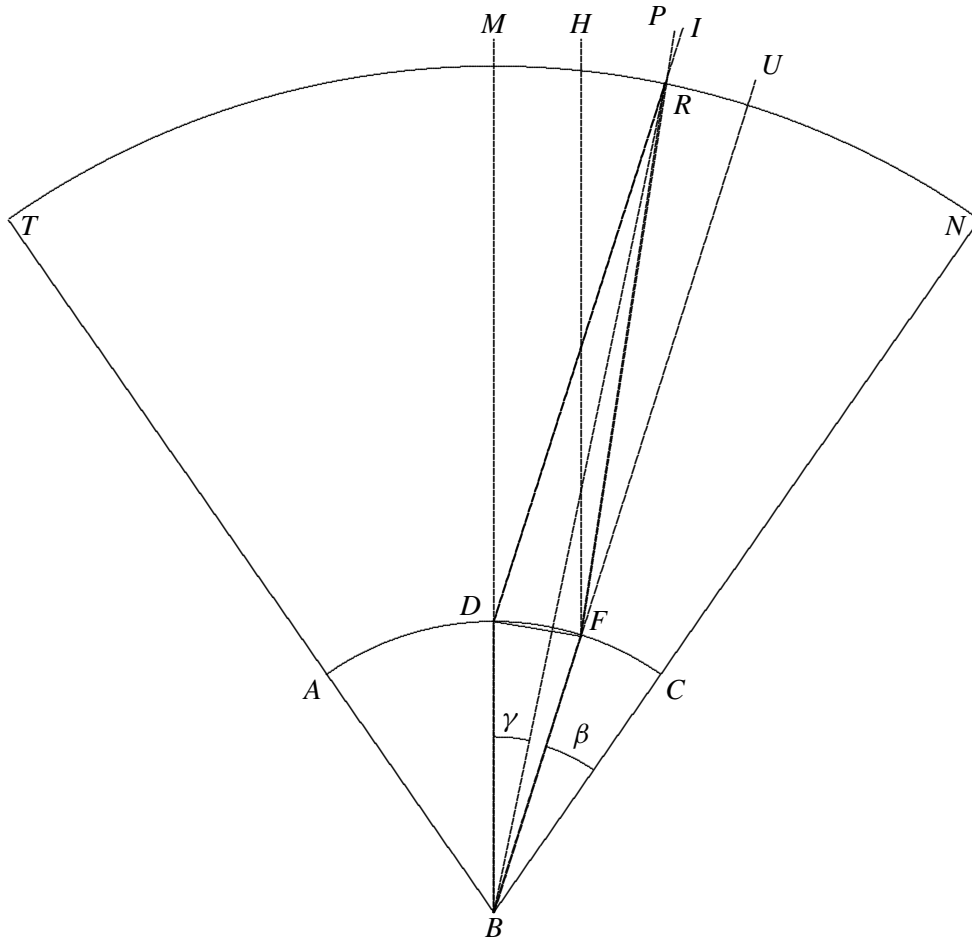


FIG. 2. A simpler picture for the proof

Addition can be performed only on like objects. Either the objects must both be sectors of the same circle, or they must both be angles. I am unable to determine which you had in mind. These two statements, and Statements 5 and 12 as well, however, can be omitted entirely. Even if Statements 8 and 11 were rendered perfectly correct, these four assertions would have no bearing on the question whether point R is a trisection point.

Actually, you have given many more lines and points than you need to perform the construction. There is no need at all to locate the point V . Once you have located R , it is an utterly trivial matter to locate V by putting the compass heel at R and its stylus at N , then drawing a circle. The other intersection of that circle with arc TN is necessarily V . I have taken the liberty of removing the unnecessary lines and points from your figure. The result is Fig. 2, in which I also drew the two lines DF and BR and marked the angles $\gamma = \angle MBR$ and $\beta = \angle UBN = \frac{1}{4}\angle ABC$. Also, it is necessary to say in Step 9 that FP intersects DI in the point R . The point R is the point through which you wish to draw the circle with B as center, meeting the extensions of BA and BC in T and N respectively.

The object is to show that $\angle RBN = \frac{1}{3}\angle ABC = \frac{4}{3}\beta$. Since $\angle RBN + \gamma = \angle NBM = 2\beta$, this means that we must prove $\gamma = 2\beta - \frac{4}{3}\beta = \frac{2}{3}\beta$. That is the *approximate* equality we shall prove.

We prove it by referring to the trapezoid $BFRD$ with parallel opposite sides BF and DR . Let us assume that the first arc drawn (arc $ADFC$) has radius r , that is, $BA = BD = BF = BC = r$. Since $\triangle DBF$ is isosceles, $\angle BDF = \angle BFD = 90^\circ - \beta/2$.

Then, since $\angle DBF = \beta$, side DF has length $2r \sin\left(\frac{\beta}{2}\right)$.

Since BF is parallel to DR , angle BDR must be $180^\circ - \beta$, and therefore angle FDR must be $(180^\circ - \beta) - (90^\circ - \beta/2) = 90^\circ - \beta/2$.

Since FH is parallel to BM , it follows that $\angle HFU = \beta$, and therefore, since FR bisects $\angle HFU$, that $\angle UFR = \beta/2$, so that $\angle BFR = 180^\circ - \beta/2$.

Since the four angles of a quadrilateral must total 360° , it now follows that $\angle DRF = 360^\circ - (180^\circ - \beta/2) - \beta - (180^\circ - \beta) = \beta/2$.

Since $\angle FDR = 90^\circ - \beta/2$, as shown above, it follows (since the sum of the angles of $\triangle FDR$ is 180°) that $\angle DFR = 90^\circ$. That is, $\triangle DFR$ is a right triangle!

This gives us the length of side DR , since $\sin(\beta/2) = \sin(\angle DRF) = \frac{DF}{DR}$, so that $\overline{DR} = \frac{\overline{DF}}{\sin(\beta/2)} = (2r \sin(\beta/2))/(\sin \beta/2) = 2r$. (Another interesting result: $\overline{DR} = 2\overline{BD}$.)

The Law of Cosines now gives us the length of BR :

$$\overline{BR}^2 = \overline{BD}^2 + \overline{DR}^2 - 2\overline{BD} \cdot \overline{DR} \cos(\angle BDR) = r^2 + 4r^2 + 4r^2 \cos \beta = 5r^2 + 4r^2 \cos \beta.$$

(Here we have used the fact that $\cos(180^\circ - \beta) = -\cos \beta$.)

Thus,

$$\overline{BR} = \sqrt{5r^2 + 4r^2 \cos \beta} = r\sqrt{5 + 4 \cos \beta}.$$

The Law of Sines now yields, at last, the expression for the angle γ :

$$\frac{\sin \gamma}{\sin(180^\circ - \beta)} = \frac{\overline{DR}}{\overline{BR}} = \frac{2r}{r\sqrt{5 + 4 \cos \beta}} = \frac{2}{\sqrt{5 + 4 \cos \beta}}.$$

Thus, since $\sin(180^\circ - \beta) = \sin \beta$, we finally have

$$\sin \gamma = \frac{2 \sin \beta}{\sqrt{5 + 4 \cos \beta}}, \text{ that is, } \gamma = \arcsin\left(\frac{2 \sin \beta}{\sqrt{5 + 4 \cos \beta}}\right).$$

Now, as stated above, the result needed for a perfect trisection is $\gamma = \frac{2}{3}\beta$. This expression is clearly *not* that, so we do not have a perfectly accurate trisection. Let us examine how good the approximation is.

By use of Taylor's series, the true expression for γ in terms of β can be expanded as a series of powers of β . That expansion, with β and γ in degrees, rather than the usual radians used in calculus, is

$$\gamma = \frac{2}{3}\beta + \frac{\pi^2}{2624400}\beta^3 + \frac{\pi^4}{1020366720000}\beta^5 + \frac{7\pi^6}{29753893552000000}\beta^7 + \dots$$

(I had a very laborious job calculating this, even with the aid of a hand-held calculator, but I'm fairly confident of my work, and I'm absolutely certain of the first coefficient.) Three things should be noted here. First, you got the main term absolutely right. That's rare. Usually when people send me their approximate trisections, the first coefficient in the expansion is a small amount off, not enough for the eye to detect when they perform the construction, but enough to show up in the coefficients of the series. Second, your construction makes the angle γ slightly too large, and hence makes the desired angle RBN slightly too small. Third, the error is remarkably tiny for such a simple construction. In practice one never needs to trisect anything but an acute angle, so β is never more than 22.5° . The main part of the error, which is the term involving β^3 , is therefore always *less than one-twentieth of a degree!* Congratulations! Further congratulations are due you, in that the angle constructed (γ) depends only on the angle ABC that you started with. Very, very often people send me trisections where the final result depends on the radius of the arc ADC that is drawn to start the construction, so that you get *different* angles, all alleged to be one-third of the original one!

Using *Mathematica*, I have graphed the angle γ yielded by your construction and superimposed it on the graph of $\frac{2}{3}\beta$. The result is shown in Fig. 3 below. As you can see, the graphs are so close together that almost no difference between them can be detected.

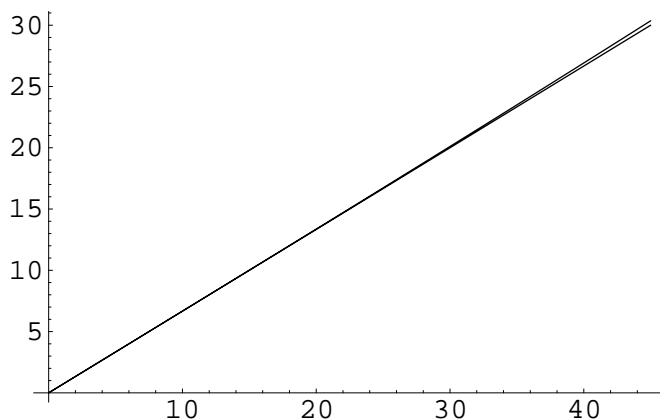


FIG. 3. Comparison of the constructed trisection with the true trisection

That's about all there is to say. If you are interested in how your approximate trisection compares with those that other people have produced, I can put you in touch with a mathematician who specializes in this subject (I don't). Just one word of warning: I doubt that your approximate trisection can be made simpler without sacrificing some of its precision or more precise without sacrificing a great deal of its simplicity. On the other hand, the facts that $\overline{DR} = 2\overline{BD}$ and that $FR \perp FD$, proved in the course of the argument, make it possible to carry out the construction with even fewer lines. (The proof, of course, does require all the lines in Fig. 2.) Here's how you could present the construction, using Fig. 4.

Draw a circle $ADCK$ with center at B , the vertex of angle ABC , the antipodal points D and K being the midpoints of the two arcs AC . Let F be the midpoint of arc CD . Draw DF . Draw FP perpendicular to DF . Using D as center and DK (the diameter of circle $ADFCCK$) as radius, draw an arc intersecting FP at point R . Then BR is the (approximate) trisector of $\angle ABC$.

Of course, to *prove* all this, you still have to go back to the argument I gave above, and the error estimate for the approximate trisection you have produced can't be made any simpler, as far as I know. But I find it fascinating that a figure with so few lines can be so close to a true trisection.

Let me close with just a gentle warning. I implore you to remember that infinite precision using these tools according to the rules is impossible. There have been some very sad cases of people who just couldn't understand that in mathematics some things really can be proved impossible, people who spend years of their lives trying to do it anyway. Proofs of *physical* impossibilities are never watertight, since the proofs are derived from *models* of the physical world they supposedly apply to, not the physical world itself. It may be inadvisable, but it's not absurd, to try to refute the laws of thermodynamics. But the world of mathematics is *not* the world of physics. The mathematical world is *identical* with the theories that model it. When those theories say something is impossible, it really is impossible. The person I could refer you to, who specializes in trisections, has a very large collection of material filed under "cranks," that is to say, the people mentioned above, who never understood, or refused to accept, the mathematical impossibility proof.

In conclusion, Mr. Clausen, I have enjoyed reading your work very much. I hope you can gain something from reading my evaluation of it. If you need further clarification, please do not hesitate to contact me. I do have one small favor to ask: I like to post at my website certain work that I do *pro bono publico*, in response to requests such as yours. If you wouldn't mind, I'd like to post this letter and the attached essay on trisections. Normally I would remove all references to your name and address from such a post, in order to protect your privacy. However, if you are pleased with your work (and you have a right to be!) and want the

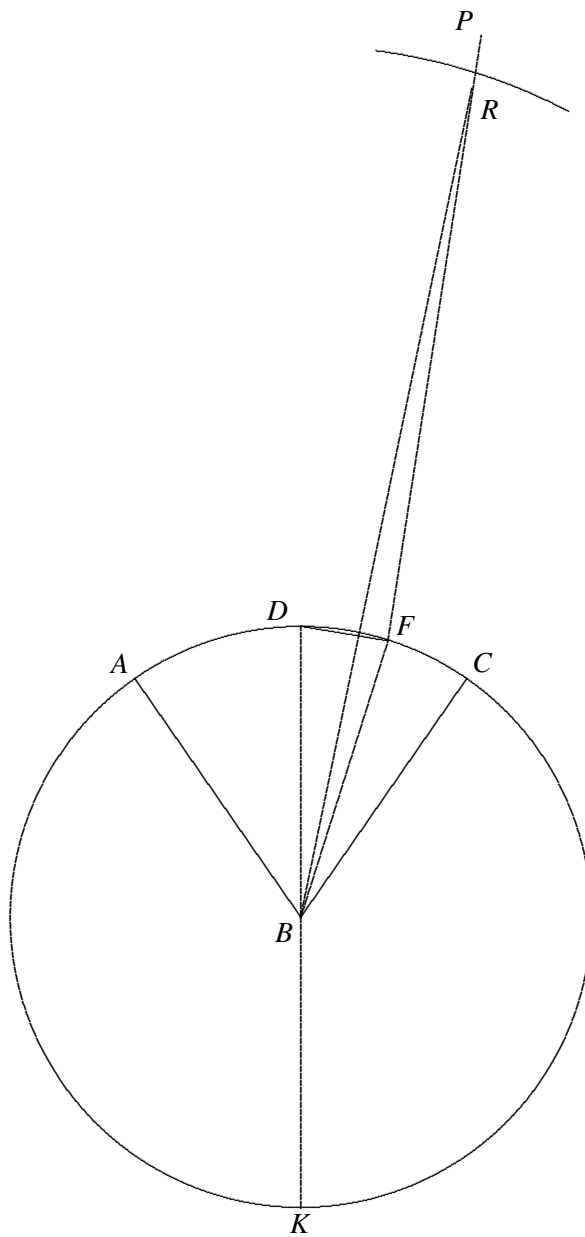


FIG. 4. Minimal presentation of the construction (without proof)

credit for having done it to be ascribed to you on the Web, I'll certainly be glad to leave your name and/or address on it. I await your permission to do this.

With best wishes,

Roger Cooke

Professor of Mathematics

email: cooke@emba.uvm.edu

url: <http://www.emba.uvm.edu/~cooke>

Angle Trisection

Pythagorean geometry, which forms most of Euclid's *Elements*, is based on the use of circles and straight lines, which can be drawn with compass and straightedge.* Other curves, such as the conic sections (ellipse, parabola, and hyperbola), were known even to the Pythagoreans, but were considered a "special topic." All the advanced curves studied by the Greeks, such as the conic sections, spirals, the cissoid of Diocles, the conchoid of Nicomedes, the quadratrix of Hippias, and others, were studied primarily in order to solve three problems that (we now know) cannot be solved using only circles and straight lines. These problems are to construct a square of area equal to a given circle, to construct a cube twice as big as a given cube, and to trisect a given angle. The first two of these problems are natural extensions of the program of the Pythagoreans, who had been able to construct a square equal to any given polygon. What could be more natural than to follow up by constructing a square equal to a given circle, then to move into three dimensions by constructing a cube equal to any given polyhedron? For these more complicated problems, the ancient geometers realized (although they could not prove they were correct) that more sophisticated curves would be needed.

The problem of trisecting an angle is part of the problem of dividing a circle into equal parts. It is a very simple construction, one learned by all students in high school geometry, to divide a line into any prescribed number of equal parts, and this can be done using only lines and circles. The situation is different with circles, however. A circle can be divided into 2, 3, 4, 5, 6, 8, 10, 12, or 17 equal parts using a straightedge and compass, but not 7, 9, or 11. It seems strange that one can construct a 60° -angle with these tools, but not a 40° -angle. That is the way things work, however. In the second century CE the astronomer Claudius Ptolemy constructed the first table of chords (equivalent to a table of the sine function), and was able to construct angles of 1.5° and 0.75° , but not 1° ; for that angle he had to use interpolation between the other two just mentioned.

If, however, one is willing to bend the rules slightly, it becomes possible to trisect an angle by drawing only straight lines and circles. Probably the most famous example of this construction comes from Archimedes' *Book of Lemmas*. The construction is illustrated in Fig. 5 and goes as follows.

Given an angle $\alpha = \angle ABC$ to be trisected, draw the semicircle $CAEF$ with B as center and BC as one extreme radius. Using a straightedge more than three times the length of BC , place two marks on it near one end, separated by a distance equal to the radius of the semicircle. Place the straightedge over the point A and move it so that one of the two marks (E) is on the semicircle and the other (D) is on the extension of CBF . Draw AD . Then the angle $\beta = \angle ADC$ is one-third of α . For, the triangles DEB and EBA are both isosceles, so that $\beta = \angle EBD$ and $\angle AEB = \angle EAB = \angle DAB$. Since α is exterior to $\triangle ADB$, we have $\alpha = \beta + \angle DAB = \beta + \angle AEB$. But angle AEB is exterior to triangle BED , and therefore $\angle AEB = 2\beta$. By substitution, then $\alpha = 3\beta$.

To achieve this construction, of course, Archimedes had to bend the rules, which specify that you can draw a line only when you can locate two points on it and a circle only when you can locate its center and one of its points. The only point on the line AD that can be located by these principles is A . The point D must be located visually, without instruments, and that is why this construction isn't a true trisection.

Although the Greeks probably realized that it was impossible to trisect every angle using only straightedge and compass, a proof of the impossibility was not completed until the early nineteenth century, when Pierre Laurent Wantzel (1814–1848) was able to use the group theory developed by Galois to give the first correct proof. We cannot develop that machinery (which requires at least a full-year course in group and field theory) in the space of this essay. However, we can at least give a heuristic argument to show why a trisection is impossible.

* A straightedge uses a physical straight line that has already been constructed as a model to draw others. That leaves open the question how the edge was made straight in the first place. A physical device—the Peaucellier linkage—that draws a true straight line, just as a compass draws a true circle, was invented only in the nineteenth century.

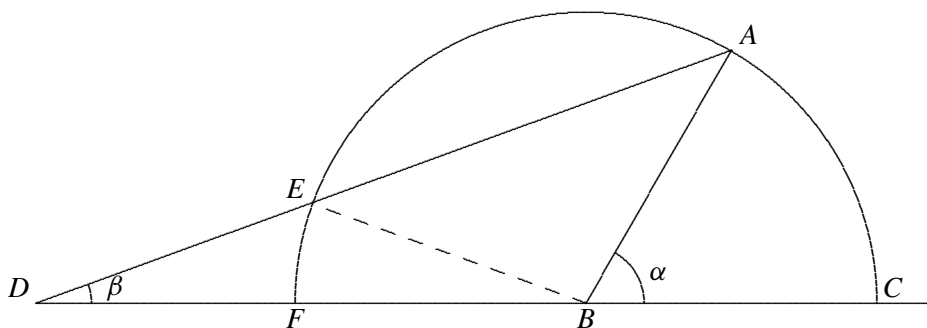


FIG. 5. ARCHIMEDES' TRISECTION OF AN ANGLE

In order to do so we need to introduce several kinds of numbers. First of all, there are *real numbers*, which can be thought of as *infinitely precise* decimal expansions. They need to be distinguished from the floating-point numbers used in everyday life and by computers, which always have limited precision. We *may* know all of the infinitely many decimal digits of a real number; we do, for instance, in the case of $\frac{1}{3} = 0.3333 \dots$. But we may not; for example, no one knows the trillionth decimal digit of π at the moment. From the real numbers, we get the *complex numbers*, which are merely sums $a + bi$, where a and b are any two real numbers and i (the electrical engineer's j) is a symbol representing a square root of -1 . The complex numbers are naturally identified with points in the plane, a (the real part) and b (the imaginary part) being the horizontal and vertical coordinates of the number $a + bi$ referred to a pair of perpendicular axes. Among the real numbers a special place is reserved for the *rational numbers*, which are ratios of integers (whole numbers). The rational numbers are essentially the only real numbers for which it is possible to know all the digits in the decimal expansion. Real numbers that are not rational are (naturally) called *irrational*. However, the classification of numbers as rational or irrational is something like classifying the universe into things that are bananas and non-bananas: One of the classes tends to be rather diverse. For that reason, we introduce one other classification. A complex number θ is *algebraic* if it satisfies a polynomial equation with integer coefficients. Every rational number $\theta = \frac{m}{n}$ is algebraic, since $n\theta - m = 0$, but there are also irrational algebraic numbers. For example, $\theta = \sqrt{2}$ satisfies $\theta^2 - 2 = 0$, the golden ratio $\varphi = \frac{1+\sqrt{5}}{2}$ satisfies $\varphi^2 - \varphi - 1 = 0$, and the complex number $\theta = 1 + \sqrt{2}i$ satisfies $\theta^2 - 2\theta + 3 = 0$. For every algebraic number θ there is one and only one polynomial $p(x) = c_0x^n + c_1x^{n-1} + \dots + c_{n-1}x + c_n$ that is of minimal degree and has the smallest possible positive leading coefficient (c_0) and is such that $p(\theta) = 0$. This polynomial is (naturally) called the *minimal polynomial* of θ ("over" the rational numbers, as mathematicians say). A number that is not algebraic is called *transcendental*, since it "transcends" the algebra of rational numbers. (But see the remarks above about the rational/irrational classification. The algebraic/transcendental dichotomy is not much better.)

With these generalities about kinds of complex numbers taken care of, we can now relate algebra to geometric constructibility. The first question one might ask is: Given two points P and Q in the plane representing the complex numbers 0 and 1, respectively, which points can be reached as the intersections of lines and circles drawn according to Euclidean rules. For example, every integer $0, \pm 1, \pm 2, \dots$ can be reached in this way. First draw the line through PQ . (It consists of all the real numbers, but so far only two of them have been "reached.") Putting the heel of the compass at 0 and drawing the circle through 1, we get a second intersection point, namely -1 . Repeating this operation using the points already located as centers, we can get all the integers. Since a line can be divided into any number of equal parts using ruler and compass, we can also get all rational numbers. Since we can also construct the perpendicular to a line at any point, we can obviously locate all complex of the form $r + si$ where r and s are rational. Moreover, we can locate the square root of any positive number r , based on the fact that it is the mean proportional

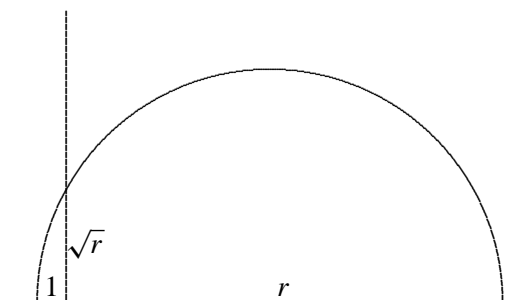


FIG. 6. CONSTRUCTION OF \sqrt{r} BY RULER AND COMPASS

between r and 1, as shown in Fig. 6.

Since the square root of any positive real number can be constructed, the formula

$$\sqrt{a + bi} = \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} + i\sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}},$$

which is correct when $b > 0$, shows that the square root of any constructible complex number can be constructed with ruler and compass.

A kind of converse of this last result also holds. Since the equation of a line is linear ($ax + by = c$) and the equation of a circle is quadratic ($(x - h)^2 + (y - k)^2 = r^2$), it follows (though the proof is long and difficult) that the *only* complex numbers one can locate in this way are algebraic numbers whose minimal polynomials are of a certain form. In particular, the degree of the minimal polynomial cannot be 3. Thus $\sqrt[3]{2}$ is *not* constructible, since its minimal polynomial over the rational numbers is $x^3 - 2$; and *that is why it is impossible to double a cube using only ruler and compass*. To double a cube means to find, with ruler and compass, the length $\sqrt[3]{2}$ times a given length.

Things are even worse in regard to squaring the circle. If we could square a circle, that is, construct the side s of a square equal to the area of a given circle πr^2 , we would have $s = r\sqrt{\pi}$, and hence we'd be able to construct the number π . But π isn't even an algebraic number, much less a constructible one. It is *transcendental*, as was proved by Ferdinand Lindemann (1852–1939), with a few corrections by Karl Weierstrass (1815–1897), in 1881.

Finally, we come to the subject of this essay, trisecting an angle. Obviously *some* angles can be trisected. One can easily trisect a right angle, for example, and hence one can trisect a 45° -degree angle. But a 60° -degree angle *cannot* be trisected. Suppose, for example, we could construct a 20° -degree angle. Imagine a right triangle ABC whose hypotenuse AB is of length 1 and with a 20° angle at A . The length of side AC would then be $\theta = \cos(20^\circ)$, so that this number would be constructible. But its minimal polynomial can be shown to be $8x^3 - 6x - 1$, which is of degree 3.

Despite the hard work of Wantzel, Lindemann, and Weierstrass in proving that these constructions are impossible, enthusiastic amateurs every year produce hundreds of new attempts on each of these problems. I would recommend to anyone who is inclined in this direction a thorough reading of *A Budget of Trisections* by Underwood Dudley. This book (ISBN 0387965688) was published by Springer-Verlag, New York, in 1987 and is now out of print, but it is in many libraries.

A person who recognizes that a mathematically exact (infinitely precise) solution following the strict Euclidean rules is impossible, but can be achieved by relaxing the rules a small amount or using curves more complicated than circles and straight lines, can spend many happy hours looking for simple and approximate solutions. This is, I think, a healthy pastime. Unfortunately too many amateurs fail to understand that the impossibility is a purely theoretical thing and that a practical solution is of no interest (that is, no mathematical journal would publish one), since dozens of such solutions were produced hundreds and even thousands of

years ago. Too many of them like to take up what appears to them to be a challenge. (“They said it couldn’t be done.”) In engineering such an attitude is admirable, since there are no absolute impossibilities in the physical world (although some things are so nearly certainly impossible that one would be a fool to attempt them, such as the construction of a perpetual motion machine). To follow this instinct in mathematics, however, is the road to madness. As Virgil said, “Facilis discensus Averni,” or, in similar language expressed by a slightly later teacher, “Wide is the gate, and broad is the way, that leadeth to destruction, and many there be that go in thereat.” Anyone who chooses to go this way should at least heed the words of a great disciple of these two: “Lasciate ogni speranza, voi ch’entrate!”