2.B. Impossibility of the classical construction problems

Ancient Greek geometry devoted a great deal of effort to construction problems like bisecting angles, dividing line segments into a given number of equal parts, finding a segment whose length is the square root of a given segment's length, constructing regular pentagons, and many others too numerous to mention. Three specific questions that resisted solutions became topics of great interest during the fifth century B. C. E. These questions have had an important impact on all of mathematics for various reasons:

1. *Angle trisection.* Given an arbitrary angle, find another one whose degree measure is one third the measure of the original angle.

2. *Circle squaring.* Given an arbitrary circle, find a square whose area is equal to that of that circle.

3. *Cube doubling.* Given an arbitrary cube, find another whose volume is twice that of the original one.

When Greek mathematicians did not find solutions to these questions using elementary tools like an **unmarked** straightedge and compass, they turned to more complicated curves and were able to solve the problems using these methods. On the other hand, there was always a desire to find constructions that only required a straightedge and compass, and Plato's insistence on the use of such methods was particularly influential.

During the nineteenth century mathematicians showed that none of the three problems above could be solved using an unmarked straightedge and a compass. These results are often misunderstood, so we shall attempt to explain what they mean and how they were obtained.

It is probably best to begin by discussing the meaning of mathematical impossibility. It does not mean that no one thus far has found an answer to a question, but instead it means that no one will ever be able to find an answer. The irrationality of $\sqrt{2}$ is a good example. One way of expressing this is to say that one cannot find positive integers p and q such that $\sqrt{2} = p/q$. This is a much stronger statement than an observation that no one has yet been able to find such a pair of numbers; in fact, the argument shows that if such a pair of numbers existed, then one would reach an absurd conclusion about the factorization of the numerators and denominators into products of primes. In particular, if m and n are reduced to least terms, the standard argument shows that one of these numbers is simultaneously odd and even.

Reformulation using real numbers and coordinates

The impossibility proofs for the three construction problems are based upon a translation of the geometric problem into algebra, and to a some extent a reformulation involving coordinate geometry. We shall be particularly interested in a set of real numbers that we shall call the *surds*. This is the smallest subset that contains all the rationals and is closed under addition, subtraction, multiplication, division and extracting square roots of positive numbers. We shall denote this set of numbers by **Surds**. Further information on surds and more details on many assertions in this document may be found in Chapter 19 of the following classic text:

> E. E. Moise. **Elemetary geometry from an advanced standpoint.** Third Edition. Addison-Wesley Advanced Book Program. *Addison-Wesley, Reading, MA*, 1990. ISBN: 0–201–50867–2.

Classical constructions by straightedge and compass start with some configuration of points and lines. In the terminology of coordinate geometry, we shall think of the points as given by ordered pairs of real numbers and the lines by equations of the form $Ax + By + C = 0$ where either $A \neq 0$ or $B \neq 0$. Let us agree to take normalized coefficients such that $A^2 + B^2 = 1$. The following result can then be established fairly directly using the fact that lines are given by first degree equations and circles by quadratic equations:

Constructibility criterion. *Suppose that one starts with a collection of points and lines such that the coordinates of the points all lie in* **Surds** *and the normalized coefficients of the lines' equations also lie in* **Surds**. *Then any point that is obtained from the original collection by a finite sequence of elementary constructions will have coordinates that also lie in* **Surds**.

Proving this result essentially requires nothing more than standard coordinate geometry, the quadratic formula, and some persistence. However, the next step is less elementary and is usually first established in graduate level algebra courses (it is also in Chapter 19 of Moise's book).

Surds as roots of rational polynomials. *If $x \in$* **Surds**, *then x is a root of a nontrivial polynomial with rational coefficients. In fact, the set of all such polynomials having x as a root consists of multiples of some minimal polynomial $m(t)$ of least degree, and this degree is a power of 2.*

We shall now explain why these two observations combine to show the impossibility of completing the classical problems by means of straightedge and compass.

In the case of doubling the cube, if the construction could be carried out then the first result would imply that the point $(2^{1/3}, 0)$ would have surd coordinates. In particular the cube root of 2 would be a surd.

In the case of squaring the circle, if the construction could be carried out then the first result would imply that the point $(\sqrt{\pi}, 0)$ would have surd coordinates. In particular the number π would be a surd.

The case of angle trisection requires a longer discussion. If the construction could be carried out for all angles, then in particular it could be carried out for a 60 degree angle. An angle of this type can be described using points whose coordinates are surds. In particular, if we take $A = (2, 0)$, $B = (0, 0)$ and $C = (2, \sqrt{3}\,)$ then $\angle\, ABC$ will be a 60 degree angle. If we could trisect this angle by means of straightedge and compass, this would mean that we could that the point $2 \cos 20°$ would be a surd.

The impossibility proof then reduces to showing the following:

THEOREM. *None of the numbers $2^{1/3}$, π, $\cos 20°$ is a surd.*

The conclusions for $2^{1/3}$ and $\cos 20°$ were established by P. Wantzel (1818–1848) in the eighteen thirties; after the publication of this work C. F. Gauss (1777–1855) stated that he had known these results over 30 years earlier, and for numerous reasons this claim is viewed as extremely credible. Proving the conclusion for π required additional input, and it was completed by F. Lindemann (1852–1939) in the eighteen eighties. During the previous decade Ch. Hermite (1822–1901) proved a corresponding result for the number

$$e \quad = \quad \sum_{n=0}^{\infty} \frac{1}{k!}$$

that plays such an important role in calculus, and Lindemann's proof relied heavily on Hermite's earlier work.

Explanation of how results are shown. The result involving π is the easiest to explain. Lindemann's result shows that π is not the root of any nontrivial polynomial; Hermite's results prove the same conclusion for e.

Discussion of the remaining two cases will be simplified with references to the following important and useful result.

FACTORIZATION PRINCIPLE FOR POLYNOMIALS. (Gauss) *Suppose that $f(x)$ is a nonconstant polynomial with integer coefficients that can be factored over the rationals into a product $f(x) = p(x) \cdot q(x)$ where p and q are polynomials of strictly lower degree. Then there are integral polynomials $p_1(x)$ and $q_1(x)$ such that $\deg p_1 = \deg p$, $\deg q_1 = \deg q$, and $f(x) = p_1(x) \cdot q_1(x)$.*

This result is often proved in upper level undergraduate algebra courses. One may view this result as a strong generalization of the fact that $\sqrt{n}$ is irrational if the positive integer n is not a perfect square for the following reasons: If $\sqrt{n}$ were a rational number r, then $x^2 - n$ would factor over the rationals as a product $(x - r)(x + r)$, and by the result of Gauss there would also be a factorization of this sort over the integers. This would mean that $x^2 - n$ had an integral root, which amounts to saying that n is a perfect square. But we were assuming that n was not a perfect square, so we have a contradiction. The contradiction arises because we assumed that $\sqrt{n}$ was rational, so the latter cannot be true.■

We now return to the issue of showing that $2^{1/3}$ and cos 20° are not surds. The result concerning the cube root of 2 is seen as follows. We know this number is a root of the rational polynomial $x^3 - 2$, and we claim the latter cannot be written as a product of two polynomials of lesser degrees with rational coefficients. If it could, then the result of Gauss would imply that one could also find such a factorization involving polynomials with integral coefficients. Since we are working with a degree 3 polynomial, one of the factors would have to be linear, and since the coefficient of x^3 is 1 this linear factor would necessarily be of the form $\pm x - a$ for some integer a. It would then follow that a would be a root of $x^3 - 2$ and that 2 would be a perfect cube over the integers. Since $1^3 = 1$ and $n^3 \geq 8$ for every integer $n > 1$, we know this is impossible. Therefore there is no way of factoring $x^3 - 2$ over the rationals into a product of two polynomials of lower degree.

On the other hand, if $2^{1/3}$ were a surd then this cubic polynomial would be a multiple of some rational polynomial whose degree is a power of 2. Since we know that $x^3 - 2$ cannot be written in this manner, it follows that the cube root of 2 cannot be a surd. In particular, this implies that one cannot solve the cube duplication problem by means of an unmarked straightedge and compass.

The result concerning cos 20° uses similar reasoning but it also requires some information on realizing this number as the root of a rational polynomial. One obtains the needed information by means of standard trigonometric identities for computing $\cos 3\theta$ in terms of $\sin\theta$ and $\cos\theta$:

$$\cos 3\theta \;=\; 4\cos^3\theta \;-\; 3\cos\theta$$

If we set $\theta = 20°$ and $\alpha = \cos 20°$, then $\cos 60° = \frac{1}{2}$ and the identity imply the following equation:

$$8\alpha^3 \;-\; 6\alpha \;-\; 1 \;=\; 0$$

Proving that α is not a surd is equivalent to proving that 2α is not a surd; this is helpful because if $\beta = 2\alpha$ then we have

$$\beta^3 \;-\; 3\beta \;-\; 1 \;=\; 0\,.$$

Once again we shall be finished if this is an irreducible polynomial over the rational numbers. As in the previous case, it is only necessary to show that the polynomial $x^3 - 3x - 1$ cannot have a linear factor with integer coefficients, and this can be checked directly as follows: If the linear integral polynomial $cx + d$ was a factor of the cubic polynomial $x^3 - 3\,x - 1$ then each coefficient would have to be either $+1$ or -1 and none of the polynomials of this form can divide $x^3 - 3\,x - 1$ (since ± 1 is not a root of the latter). Therefore we conclude that cos 20° cannot be a surd, and this shows that one cannot trisect a 60 degree angle by means of unmarked straightedge and compass.

It is important to note that **some** angles can be trisected using straightedge and compass. For example, one can trisect a 45 degree angle because it is possible to construct a 45 degree angle, a 60 degree angle, and an angle whose degree measure is the difference between those of the first two — a difference that is merely 15 degrees. The point of the problem is to be able to carry out the construction for **all** angles.

Comment on the regular pentagon

We have mentioned that a regular pentagon can be constructed using straightedge and compass. In particular, this implies that the sine and cosine of $36° = \frac{1}{5} \cdot 180°$ must be a surd. In fact these numbers are closely related to the roots of the quadratic equation $x^2 + x - 1 = 0$, which are given by

$$-\frac{1}{2} \pm \frac{\sqrt{5}}{2}$$

that have had mystical attractions ever since the time of the ancient Greeks (the recent book on the Da Vinci Code is a very current example).

In order to verify that $\cos 36°$ is a surd one must begin with the appropriate formula expressing $\cos\, 5\,\theta$ in terms of sin *theta* and $\cos\,\theta$. This can be derived by direct calculation (a quicker alternative using complex numbers would be to note that $\cos\, 5\,\theta$ is the real part of

$$\exp(5\,i\,\theta) \quad = \quad \exp(i\,\theta)^5 \quad = \quad (\cos\,\theta \;+\; i\,\sin\,\theta)^5$$

and evaluate it in terms of $\sin\,\theta$ and $\cos\,\theta$ using these equations). Here is the end result:

$$\cos\, 5\,\theta \quad = \quad 16\,\cos^5\,\theta \;-\; 20\,\cos^3\,\theta \;+\; 5\,\cos\,\theta$$

If we set $\theta = 18°$, so that $2\,\theta = 36°$ and $5\,\theta = 90°$, then it follows that $\alpha = \cos\, 18°$ satisfies the equation

$$16\,\alpha^5 \;-\; 20\,\alpha^3 \;+\; 5\,\alpha \quad = \quad 0$$

and since $\alpha \neq 0$ we conclude that $\beta = 2\,\alpha$ satisfies the equation

$$\beta^4 \;-\; 5\,\beta^2 \;+\; 5 \quad = \quad 0\;.$$

The Quadratic Formula then implies that

$$\beta^2 \quad = \quad \frac{5}{2} \pm \frac{\sqrt{5}}{2}$$

and since $\beta = 2\,\cos\, 18°$ it follows that β^2 must be the root with the positive sign. Therefore we conclude that

$$\cos^2\, 18° \quad = \quad \frac{\beta^2}{4} \quad = \quad \frac{5 + \sqrt{5}}{8}$$

and using the identity $\cos\, 2\,\theta = 2\,\cos^2\,\theta - 1$ we obtain the following formula for $\cos 36°$:

$$\cos\, 36° \quad = \quad \frac{1\;+\;\sqrt{5}}{4}$$

In other words, the cosine of 36° is half the larger root of the equation $x^2 + x - 1 = 0$.■

Most numbers of the form $\cos(180°/n)$ are not expressible as surds. In fact, results of Gauss show this only happens if n is product $2^k p_1 \cdots p_m$ where each p_j is an odd prime of the form $2^{s_j} + 1$. Clearly the primes 3 and 5 have this form, and the next such prime is 17; more generally, if $2^s + 1$ is a prime then s must be a power of 2, and thus the next prime after 17 in this list is 257. The result of Gauss has the following geometric implication: Given values of n are the only ones for which a regular n-gon can be constructed using an unmarked straightedge and compass.

Common misconceptions and mistakes

Over the 24 centuries since the classical construction problems first attracted attention in ancient Greece, many people have incorrectly claimed or believed that they had solved one or more of these problems. Others have felt it is short-sighted or presumptuous for mathematicians to state that the problems are impossible to solve. Therefore it seems worthwhile to discuss some points that arise repeatedly in the discussion of these "impossible" construction problems.

Probably the most important source of misunderstanding is the insistence that the construction be carried out using only an unmarked straightedge and a collapsible compass. There are many solutions of each construction problem if one relaxes these requirements even slightly. For example, if one merely allows the use of a straightedge with two markings (an extremely primitive ruler!), then one can construct angle trisectors, and in fact this was known to Archimedes in the Third Century B. C. E.

Before discussing other misconceptions associated to the classical construction problems, some comments on the straightedge and compass limitation may be worthwhile. When Plato formulated the restriction to unmarked straightedges and compasses, he knew that solutions were possible if one broadened the list of tools that could be used, but he felt that the introduction of numerous mechanical devices compromised the purity of mathematics. Even if one rejects this view, there are often good reasons to search for ways of solving mathematical problems that are somehow restricted. Some are practical (one does not always have unlimited resources), others are conceptual (*e.g.*, the principle of Ockham's razor: One should not increase, beyond what is necessary, the number of entities required to explain anything), and still others are based upon past experience (sometimes the restricted methods turn out to be useful for other, unanticipated purposes — this is an aspect of what some call the "unreasonable effectiveness of mathematics").

Another frequent misunderstanding arises from the concept of mathematical impossibility. Throughout history people have said that certain things were impossible and have been proven wrong. Why should the impossibility statements for the three classical construction problems be any more reliable than, say, earlier assertions that objects like airplanes could never be constructed?

Part of the answer goes back to the strict rules for construction by straightedge and compass. To illustrate the difference, it is good to consider one prominent mathematician's claim in the late nineteenth century that airplanes could never be built. The work assumed that the only way to generate sufficient power would be to use a steam engine, which is extremely heavy. In fact, the important breakthrough in powered flight was the idea of using petroleum fueled engines to produce the power needed to get off the ground; no one has ever built an airplane powered by a steam engine, but for our purposes the value of this statement is entirely theoretical.

Another response is that one wants exact constructions; from the perspective of the problem, it is not enough to be able to do something within an arbitrarily small degree of accuracy. It is possible to solve **all** these problems to an arbitrarily high degree of approximation by unmarked straightedge and compass. For example, one can see this by considering the decimal expansions of real numbers like the cube root of 2, the cosine of 20°, and the number $\sqrt{\pi}$. In each case one can obtain an arbitrarily good approximation by a finite decimal fraction (which may have an astronomical number of terms), and there is no problem constructing segments of length $2^{1/3}$ and $\sqrt{\pi}$ from a segment of unit length by the classical methods. If we want to trisect an arbitrary angle to any desired degree of accuracy, one systematic way of proceeding is to use the geometric series

$$\frac{1}{3} \quad = \quad \sum_{m=1}^{\infty} \frac{1}{4^m}$$

and the fact that one can bisect angles by straightedge and compass. Repeated bisections of an angle $\angle ABC$ with measure θ will yield a sequece of angles $\angle ABC_n$ whose measures are equal to the first n terms of the infinite sum, and these measures approximate $\frac{1}{3}\theta$ to an arbitrarily high degree of accuracy if we take n to be sufficiently large.

A third source of difficulties involves the precise meaning of mathematical impossibility. As noted before, it does not mean that people have simply given up on a problem but maybe some future genius will succeed where others have failed. Instead, it means that if the problem had a solution, then one could also do something else that contradicts known facts. The irrationality proof for $\sqrt{2}$ discussed earlier in this document is an excellent example.

The preceding discussion has an important consequence: *If anyone produces an argument purporting to solve one of the classical construction problems by unmarked straightedge and compass, then there must be a mistake somewhere in the argument.* However, finding these mistakes in any given argument can be extremely difficult. A very simple example of a fallacious argument is discussed in another supplement to the main unit.

Some online references

Further discussion of mathematical impossibility and trisections is given in the first online document listed below. The second one by the same author considers scientific fallacies more generally, and the third discusses another relatively classical construction problem from late Greek and Arab mathematics (the Al-Hazen Billiard Problem) whose impossibility was not proven until fairly recently.

`http://www.uwgb.edu/dutchs/PSEUDOSC/trisect.HTM`

`http://www.uwgb.edu/dutchs/pscindex.htm`

`http://mathworld.wolfram.com/AlHazensBilliardProblem.html`