Reflections on Reflection in a Spherical Mirror

Author(s): Peter M. Neumann

Source: *The American Mathematical Monthly*, Jun. - Jul., 1998, Vol. 105, No. 6 (Jun. - Jul., 1998), pp. 523-528

Published by: Taylor & Francis, Ltd. on behalf of the Mathematical Association of America

Stable URL: https://www.jstor.org/stable/2589403

Reflections on Reflection in a Spherical Mirror

Peter M. Neumann

1. ALHAZEN'S PROBLEM. There is an old question in optics that has been called *Alhazen's Problem*: given a spherical mirror and points A, B in space, how can a point P on the mirror be found, where a ray of light is reflected from A to B? Since P must lie in the plane containing A, B and the centre of the sphere, this is really a two-dimensional problem: given a circular mirror and points A, B in its plane, find the point on the circle where a ray of light is reflected by the mirror from A to B. It is this version of the problem that we shall discuss here. Dörrie **[1]** refers to it as *Alhazen's Billiard Problem* for reasons which are not far to seek. The name Alhazen honours an Arab scholar Ibn al-Haytham who flourished 1000 years ago. The problem itself can be traced further back, at least to Ptolemy's *Optics* written some time around AD 150. A charming account, full of interesting historical and bibliographical pointers, has been published by John D. Smith **[8]**.

If the mirror is a complete sphere or circle then the problem makes physical sense only if A, B lie on the same side—both inside as in Figure 1, or both outside and visible to each other. Typically there are two or four reflection points (Drexler & Gander **[2]** discuss this further). Various methods for finding them are described in **[1]**, **[8]** and references cited there. Most describe P as the point of intersection of the mirror with an explicitly given hyperbola or cubic curve. It seems, however, that the question whether a reflection point can be found by classical Ruler & Compass methods has not been answered before—perhaps it has not even been asked.

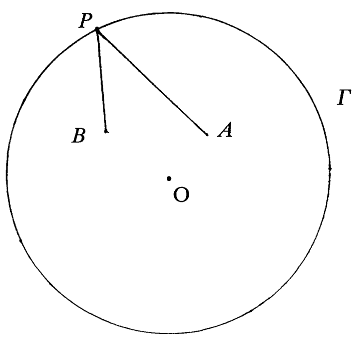

Figure 1

Theorem. *Let* Γ *be a circle and let* A, B *be points in its plane. In general there is no ruler and compass construction to find a point* P *on* Γ *at which a ray of light from* A *is reflected to* B.

To understand this theorem one first needs to know precisely what is meant by a ruler and compass construction. Although it is widely known and loved the theory is sketched in the next section. But only sketched: for details the reader is referred to any of [**1**], [**3**], [**4**], [**5**], [**6**], [**9**]. One also needs to know what is intended by those cautious words 'in general'. In the first instance they simply mean that *there exist* points A and B such that P cannot be found by a ruler and compass construction whose initial data consist of Γ, A, B. Of course there are circumstances under which there is such a construction—for example, if A, B lie equidistant from the centre of the circle then P can be obtained as one of the points where Γ meets the perpendicular bisector of AB. Going considerably deeper, the words 'in general' mean that only in very special cases does there exist a ruler and compass construction to find P from the data. We shall return to this point at the end of the paper.

2. RULER AND COMPASS CONSTRUCTIBILITY. Briefly, a ruler and compass construction is a process which, in our context, begins with Γ, A, B as known. Each step of the construction consists of one of the following: drawing the straight line through any chosen pair of already known points; drawing the circle whose centre is a known point and which passes through another known point; identifying a new point as the intersection of two known lines, a known line and a known circle, or two known circles. To prove theorems about constructibility we move from geometry to the algebra of cartesian coordinates and use a theory that is nowadays taught as part of second- or third-year university courses in abstract algebra, namely the theory of fields [**3**], [**6**], [**9**]. What it tells us in the context of Alhazen's Problem is this. Take the centre of Γ to be $(0, 0)$, take its radius to be r, and take A, B to be the points (a_1, a_2) and (b_1, b_2) respectively. Call a real number *constructible* if, starting from the numbers r, a_1, a_2, b_1, b_2, it can be obtained by a finite number of the operations $+$, $-$, $\times$, $\div$, $\sqrt{\ }$. It is important that a number is to be called constructible only if it can be obtained by the usual arithmetical operations together with extraction of *square* roots. Cube roots, for example, are not available. The general theory tells us that if, starting from the initial data Γ, A, B there is a ruler and compass construction for the point (x, y), then x and y must be constructible in the sense just defined. It also tells us the converse, but that is not relevant here.

As has already been mentioned, we analyse constructibility by using the theory of fields. For us it is not necessary to consider the general abstract theory. Subfields of the field of all complex numbers suffice. Thus we may take a field to be a set of complex numbers that contains 0 and 1 and is closed under the usual arithmetical operations $+$, $-$, $\times$, $\div$. Recall that if F, K are fields and $F \subseteq K$ then K may be thought of as a vector space over F. The *degree* $|K : F|$ of K over F is defined to be $\dim_F(K)$, the dimension of K as a vector space over F. One of the most important facts about degree is that it behaves multiplicatively. The theorem ([**3**, Theorem 5.1.1], [**6**, Lemma 31], [**9**, Theorem 4.2]) which many students nowadays call the "Tower Theorem" states that if F, K, L are fields and $F \subseteq K \subseteq L$ then $|L : F| = |L : K| \times |K : F|$.

Define F_0 to be the field $\mathbb{Q}(r, a_1, a_2, b_1, b_2)$ of all those real numbers that can be obtained from the radius of Γ and the coefficients of A and B by the rational operations $+$, $-$, $\times$, $\div$. We shall call a field F an *iterated quadratic extension* of F_0 if there is a sequence of fields $F_0 \subset F_1 \subset \cdots \subset F_k$ such that $F_k = F$ and $|F_{i+1} : F_i| = 2$ for $0 \leq i \leq k - 1$. Using induction and the "Tower Theorem" one sees that $|F_k : F_0| = 2^k$, that is, the degree of an iterated quadratic extension is a power of 2. The relevance of this is that if a point (x, y) is ruler and compass constructible from Γ, A, B then there must be an iterated quadratic extension F of F_0 such that $x, y \in F$; compare [**3**, p. 230], [**6**, Appendix 3], or [**9**, Theorem 5.2].

3. PROOF OF THE THEOREM. We take Γ to have radius 1 so that its equation is $x^2 + y^2 = 1$, we take $A := (\frac{1}{6}, \frac{1}{6})$, $B := (-\frac{1}{2}, \frac{1}{2})$, and $P := (x_1, y_1)$.

Lemma 1. $x_1^4 - 2x_1^3 + 4x_1^2 + 2x_1 - 1 = 0$.

Proof: Let's begin with quite a general treatment following that of John Smith in [**8**]. Thus A, B, P are represented by complex numbers a, b, z respectively. The condition that angles APO and OPB are equal (where O is the centre of Γ) is that $\arg((a-z)/z) = \arg(z/(z-b))$. This requires that $(a-z)(b-z)/z^2$ be real, that is, that $(a-z)(b-z)/z^2 = (\bar{a}-\bar{z})(\bar{b}-\bar{z})/\bar{z}^2$, where bar denotes complex conjugation. Rearrangement of this equation yields the condition $\bar{a}\bar{b}z^2 - ab\bar{z}^2 = ((\bar{a}+\bar{b})z - (a+b)\bar{z})z\bar{z}$. Since $z\bar{z} = 1$ this reduces to

$$\bar{a}\bar{b}z^2 - ab\bar{z}^2 = (\bar{a}+\bar{b})z - (a+b)\bar{z}, \tag{1}$$

which, after a factor $\sqrt{-1}$ has been cancelled from both sides, becomes a real quadratic equation in the coordinates x, y of P. Eliminating y between this and $x^2 + y^2 = 1$ we obtain a quartic equation for x.

For our special configuration, since $a = \frac{1}{6}(1 + \sqrt{-1})$, $b = \frac{1}{2}(-1 + \sqrt{-1})$, and P is represented by $x_1 + y_1\sqrt{-1}$), the Equation (1) becomes $x_1 y_1 - 2x_1 - y_1 = 0$. Eliminating y_1 from the equation $x_1^2 + y_1^2 = 1$ we find that $x_1^4 - 2x_1^3 + 4x_1^2 + 2x_1 - 1 = 0$, as the lemma states.

It will turn out to be helpful to know that this equation is irreducible over $\mathbb{Q}$, that is to say, that the polynomial cannot be factorised in a non-trivial way into factors with rational coefficients.

Lemma 2. *Let* $f(x) := x^4 - 2x^3 + 4x^2 + 2x - 1$. *Then* $f(x)$ *is irreducible in the polynomial ring* $\mathbb{Q}[x]$. *Moreover, two of the roots of the equation* $f(x) = 0$ *are real and the others form a complex conjugate pair.*

Proof: By Gauss's Lemma (see [**3**, p. 160], [**6**, Theorem 23], or [**9**, p. 19]), if $f(x)$ were reducible in $\mathbb{Q}[x]$ then it would have factors with integer coefficients. Clearly the leading coefficients can be taken to be 1, and the constant coefficients are ± 1. Since $f(1) = f(-1) = 4$, neither $(x-1)$ nor $(x+1)$ is a factor. Therefore the only possible factorisation is of the form

$$x^4 - 2x^3 + 4x^2 + 2x - 1 = (x^2 + cx + 1)(x^2 + dx - 1)$$

where $c, d \in \mathbb{Z}$. From the coefficient of x^3 we see that $c + d = -2$; similarly, from the coefficient of x we find that $-c + d = 2$. Therefore c has to be -2 and d has to be 0. But then $f(1) = 0$, which is not true. Thus $f(x)$ is irreducible.

Since $f(-1) > 0$, $f(0) < 0$, and $f(1) > 0$ there are at least two real roots, one between -1 and 0 and another between 0 and 1. Now $f''(x) = 12x^2 - 12x + 8 = 3(2x-1)^2 + 5$, and so $f''(x) \geq 5$ for all real values of x. Therefore $f'(x)$ is strictly increasing, and the equation $f'(x) = 0$ has at most one real root. Between any two roots of $f(x) = 0$ there lies a root of $f'(x) = 0$, and it follows that $f(x) = 0$ has at most two real roots. Thus it has exactly two real roots. Its non-real roots form a complex conjugate pair because the coefficients of the equation are real.

The nature of the roots is unimportant for our main argument (it will be exploited in Section 4), but the significance of irreducibility is as follows. Let

x_1, x_2, x_3, x_4 be the four roots of $f(x) = 0$. Since the equation is irreducible, one of them lies in an iterated quadratic extension of $\mathbb{Q}$ if and only if they all do (see [**3**, Theorem 5.3.4], [**6**, Theorem B3], or [**9**, Prop. 10.2]). It follows that if there were a ruler and compass construction for P then the splitting field of $f(x)$, that is, the field $\mathbb{Q}(x_1, x_2, x_3, x_4)$ generated by all four roots, would have to be an iterated quadratic extension of $\mathbb{Q}$. In particular, its degree would be a power of 2 (in fact, its degree would have to be 4 or 8—but we do not need to know this). Our strategy is to prove that this is not so by showing that the degree $|\mathbb{Q}(x_1, x_2, x_3, x_4) : \mathbb{Q}|$ is divisible by 3.

Lemma 3. *Define* $t_1 := x_1x_2 + x_3x_4$, $t_2 := x_1x_3 + x_2x_4$, $t_3 := x_1x_4 + x_2x_3$, *and* $g(t) := t^3 - 4t^2 - 16$. *Then* t_1, t_2, t_3 *are the roots of the equation* $g(t) = 0$. *Moreover, this equation is irreducible.*

Proof: To find the cubic polynomial whose roots are t_1, t_2, t_3 we calculate their elementary symmetric functions:

$$\sum t_i = \sum x_i x_j = 4;$$

$$\sum t_i t_j = \sum x_i^2 x_j x_k = \left(\sum x_i\right)\left(\sum x_i x_k x_k\right) - 4x_1x_2x_3x_4 = 0;$$

$$t_1 t_2 t_3 = \sum x_i^3 x_j x_k x_l + \sum x_i^2 x_j^2 x_k^2 = \cdots = 16.$$

Therefore t_1, t_2, t_3 are the roots of the equation $t^3 - 4t^2 - 16 = 0$, as the first assertion of the lemma states. Setting $u := \frac{1}{2}t$ we find that $u^3 - 2u^2 - 2 = 0$. If the polynomial $u^3 - 2u^2 - 2$ were reducible in the polynomial ring $\mathbb{Q}[u]$ then by Gauss's Lemma it would have to be factorisable in $\mathbb{Z}[u]$. Since one of its factors would have to be linear, it would have to be divisible by $u \pm 1$ or $u \pm 2$. But none of $1, -1, 2, -2$ is a root. Therefore $u^3 - 2u^2 - 2$ is irreducible in $\mathbb{Q}[u]$, and it follows immediately that $t^3 - 4t^2 - 16$ is irreducible in $\mathbb{Q}[t]$, as required.

The proof of the theorem may now be completed in a few lines. It follows from Lemma 3 that $\mathbb{Q}(t_1)$, the field generated by t_1, has degree 3 over $\mathbb{Q}$ (see [**3**, §5.3], [**6**, Theorem 28], or [**9**, Prop. 4.3]). Then, since $\mathbb{Q} \subseteq \mathbb{Q}[t_1] \subseteq \mathbb{Q}[x_1, x_2, x_3, x_4]$, by the "Tower Theorem", $|\mathbb{Q}(x_1, x_2, x_3, x_4) : \mathbb{Q}|$ is divisible by 3. Thus the splitting field of $x^4 - 2x^3 + 4x^2 + 2x - 1$ is not an iterated quadratic extension of $\mathbb{Q}$, and it follows (see the discussion preceding Lemma 3) that with our initial data the reflection point P *cannot* be constructed by ruler and compass.

4. COMMENTARY. NOTE 1. The classical proofs of the impossibility of squaring circles, duplicating cubes, and trisecting angles do not really require any Galois theory. They require only the beginnings of the theory of fields. Thus, for example, they are treated in Chapters 5 and 6 of [**9**], a book that has 19 chapters in all. By way of comparison, note that in the paragraph preceding Lemma 3 a result that appears in Chapter 10 of [**9**] has been used. It may well be that the argument can be simplified. Nevertheless, Alhazen's Problem seems to lie just a little deeper than the classical problems.

NOTE 2. The first section of this paper ends with a promise of further discussion of the words 'in general' that occur in the statement of the theorem. To keep that promise we briefly sketch some rather more sophisticated ideas.

Since the polynomial $x^4 - 2x^3 + 4x^2 + 2x - 1$ is irreducible over $\mathbb{Q}$ the degree of its splitting field is divisible by 4. We have shown that this degree is also divisible by 3, and therefore it is a multiple of 12. The Galois group of the polynomial is a subgroup of Sym (4), the group of all permutations of the four roots x_1, x_2, x_3, x_4, and, by the so-called Fundamental Theorem of Galois Theory, its order is the same as the degree of the splitting field. Therefore the order of the Galois group is divisible by 12. By the second assertion of Lemma 2, complex conjugation, which is certainly a member of the Galois group, acts as a transposition fixing two of the roots and interchanging the other two. It follows immediately that the Galois group is in fact Sym (4).

Consider now the case where A, B are points whose coefficients a_1, a_2, b_1, b_2 are algebraically independent transcendental numbers, and let G be the Galois group of the quartic polynomial for the x-coefficient of P. A tool known as 'specialisation' in algebraic geometry tells us that there is a homomorphism from G onto the Galois group that arises in the case treated in the main proof, where the points are $(\frac{1}{6}, \frac{1}{6})$ and $(-\frac{1}{2}, \frac{1}{2})$. Consequently G must also be Sym (4), and so also for this 'totally transcendental' case Alhazen's problem is insoluble by ruler and compass.

This means, of course, that if Alhazen's Problem for the points A, B is soluble by ruler and compass then the coefficients cannot be algebraically independent. Now four real numbers r_1, r_2, r_3, r_4 are not algebraically independent transcendentals if and only if there is a polynomial equation $p(r_1, r_2, r_3, r_4) = 0$ with integer coefficients. The zero-set of such an equation, being a 3-dimensional hyper-surface in $\mathbb{R}^4$, is a null set in the sense of Lebesgue measure theory. There are only countably many polynomial equations with integer coefficients and the union of a countable family of null sets is a null set. It follows that the set of pairs of points A, B for which Alhazen's Problem does have a ruler and compass solution is a null set. In other words, if A, B are chosen randomly then the probability that the reflection point P can be constructed is 0. In fact, from Hilbert's Irreducibility Theorem [**7**, Ch. 3] one can deduce that even if the coordinates a_1, a_2, b_1, b_2 of the points A, B are restricted by, say, being required to be rational numbers, they have to satisfy very strong conditions if the Galois group of the equation is to be a proper subgroup of Sym (4), as it must be if P is to be constructible with ruler and compass. That is what was meant by the remark that only in very special cases does there exist a ruler and compass construction for P.

ACKNOWLEDGMENTS. I am very grateful to my wife Sylvia, to Anne Watson and to John Smith for introducing me to Alhazen's Problem, for useful and encouraging discussions, and for constructive comment on the first draft of this paper.

REFERENCES

1. Heinrich Dörrie, *A hundred great problems of elementary mathematics: their history and solution.* (Translated by David Antin) Dover, New York, 1965.
2. Michael Drexler and Martin J. Gander, Circular billiard, *SIAM Review*, to appear.
3. I. N. Herstein, *Topics in algebra.* (Second edition) Wiley, New York, 1975.
4. Hilda P. Hudson, *Ruler & Compasses.* Longmans Green, London, 1916 (reprinted in *Squaring the circle and other monographs*, Chelsea, New York, 1953).
5. Felix Klein, *Famous problems of elementary geometry.* (Translated by W. W. Beman and D. E. Smith; Second edition with notes by R. C. Archibald) Stechert & Co., New York, 1930 (reprinted in *Famous problems and other monographs*, Chelsea, New York, 1955).
6. Joseph Rotman, *Galois theory.* Springer-Verlag, New York, 1990.

7. Jean-Pierre Serre, *Topics in Galois theory*. Research notes in Mathematics, Vol. 1. Jones & Bartlett, Boston, 1992.
8. John D. Smith, The remarkable Ibn al-Haytham, *Math. Gazette*, 76 (1992), 189–198.
9. Ian Stewart, *Galois theory*. (Second edition) Chapman & Hall, London, 1989.

PETER NEUMANN was born in Oxford in 1940 and returned there in 1959 to embark on training for a career in the family mathematical business. He holds BA, DPhil, and DSc degrees from Oxford University and has been a Fellow of The Queen's College for over one-third of a century. His teaching ranges widely over pure mathematics, but his research has been concentrated mainly in group theory. When the exigencies of academic work permit he enjoys playing with any or all of his eight grandchildren.
The Queen's College, Oxford OX1 4AW, England
Peter.Neumann@queens.ox.ac.uk

DePugh recalls a sermon he once heard at a church-ful of German Mysticks. "It might have been a lecture in Mathematics. Hell, beneath our feet, bounded—Heaven, above our pates, unbounded. Hell a collapsing Sphere, Heaven an expanding one. The enclosure of Punishment, the release of Salvation. Sin leading us as naturally to Hell and Compression, as doth Grace to Heaven, and Rarefaction. Thus—"

Murmurs of, " 'Thus'?"

"—may each point of Heaven be mapp'd, or projected, upon each point of Hell, and vice versa. And what intercepts the Projection, about mid-way (reckon'd logarithmickally) between? why, this very Earth, and our lives here upon it. We only think we occupy a solid, Brick-and-Timber City,—in Reality, we live upon a Map. Perhaps even our Lives are but representations of Truer Lives, pursued above and below, as to Philadelphia correspond both a vast Heavenly City, and a crowded niche of Hell, each element of one faithfully mirror'd in the others."

"There are a Mason and Dixon in Hell, you mean?" inquires Ethelmer, "attempting eternally to draw a perfect Arc of Considerably Lesser Circle?"

"Impossible," ventures the Rev^{d}. "For is Hell, by this Scheme, not a Point, without Dimension?"

"Indeed. Yet, suppose Hell to be *almost* a Point," argues the doughty DePugh, already Wrangler material, "—they would then be inscribing their Line eternal, upon the inner surface of the smallest possible Spheroid that can be imagin'd, and then some."

"More of these . . . ," Ethelmer pretending to struggle for a Modifier that will not offend the Company, "*curious* Infinitesimals, Cousin.—The Masters at *my* Purgatory are bewitch'd by the confounded things. Epsilons, usually. Miserable little,"—Squiggling in the air, "sort of things. Eh?"

"See them often," sighs DePugh. "this Session more than ever."

"What puzzles me, DeP., is that if the volume of Hell may be taken as small as you like, yet the Souls therein must be ever smaller, mustn't they,—there being, by now, easily millions there?"

"Aye, assuming one of the terms of Damnation be to keep just enough of one's size and weight to feel oppressively crowded,—taking as a model the old Black Hole of Calcutta, if you like,—the Soul's Volume must be an Epsilon one degree smaller,—a Sub-epsilon."

" 'The Epsilonicks of Damnation.' Well, well. There's my next Sermon," remarks Uncle Wicks.

"I observe," Tenebræ transform'd by the pale taper-light to some beautiful Needlewoman in an old Painting, "of both of you, that your fascination with Hell is match'd only by your disregard of Heaven. Why should the Surveyors not be found there Above,"—gesturing with her Needle, a Curve-Ensemble of Embroidery Floss, of a nearly invisible gray, trailing after, in currents rais'd by Talking, Pacing, Fanning, Approaching, Withdrawing, and whatever else there be to indoor Life,—"drifting about, chaining the endless airy Leagues, themselves approaching a condition of pure Geometry?"

"Tho' for symmetry's sake," interposes De Pugh, "we ought to say, '*almost* endless.' "

"Why," whispers Bræ, "whoever said anything had to be symmetrickal?" The Lads, puzzl'd, exchange a quick Look.

Thomas Pynchon, *Mason and Dixon*, Henry Holt and Company, Inc., 1997, p. 482

Contributed by William Mueller, University of Arizona, Tucson, AZ