An Elementary Proof of Marden's Theorem

Dan Kalman

What I call Marden's Theorem is one of my favorite results in mathematics. It establishes a fantastic relation between the geometry of plane figures and the relative positions of roots of a polynomial and its derivative. Although it has a much more general statement, here is the version that I like best:

Marden's Theorem. *Let $p(z)$ be a third-degree polynomial with complex coefficients, and whose roots z_1, z_2, and z_3 are noncollinear points in the complex plane. Let T be the triangle with vertices at z_1, z_2, and z_3. There is a unique ellipse inscribed in T and tangent to the sides at their midpoints. The foci of this ellipse are the roots of $p'(z)$.*

I call this Marden's Theorem because I first read it in M. Marden's wonderful book [**6**]. But this material appeared previously in Marden's earlier paper [**5**]. In both sources Marden attributes the theorem to Siebeck, citing a paper from 1864 [**8**]. Indeed, Marden reports appearances of various versions of the theorem in nine papers spanning the period from 1864 to 1928. Of particular interest in what follows below is an 1892 paper by Maxime Bôcher [**1**].

In his presentation Marden states the theorem in a more general form than given above, corresponding to the logarithmic derivative of a product $(z - z_1)^{m_1}(z - z_2)^{m_2}(z - z_3)^{m_3}$ where the only restriction on the exponents m_j is that they be nonzero, and with a general conic section taking the place of the ellipse. For this discovery he credits Linfield [**4**], who obtained it as a corollary to an even more general result "established by the use of line coordinates and polar forms." Marden asserts the desirability of a more elementary proof, and proceeds to give one based on the optical properties of conic sections.

Interestingly, Marden's proof, which appears in basically the same form in both his paper and his book, is incomplete for reasons that will be made clear below. A closely related argument in Bôcher's paper is also incomplete, although in a different way. By combining the two arguments, a complete proof of Marden's Theorem is obtained. Moreover, the proof is completely elementary, requiring very little beyond standard topics from undergraduate mathematics. To be honest, there are rather a lot of these topics required, spanning analytic geometry, linear algebra, complex analysis, calculus, and properties of polynomials. To me, the way all of these topics weave together is part of the charm of the theorem, and presenting the proof is the primary motivation for this paper. In an on-line paper [**3**] a more completely self-contained exposition is provided, with hypertext links to discussions of many of the necessary background topics, as well as animated graphics dramatizing some of the geometric ideas of the proof.

Before proceeding further, some additional observations may be illuminating. First, it is possible that the unique inscribed ellipse mentioned in the theorem is actually a circle. In this case the foci coincide, indicating that $p'(z)$ has a double root. This case can only occur if the circumscribing triangle is equilateral. In fact, this special case is easy to verify by assuming that $p'(z)$ has a double root, and deducing the form of p. The roots of p are then seen to be vertices of an equilateral triangle centered at the repeated root of p'.

Second, the inscribed ellipse mentioned in the theorem is worthy of further consideration. It is sometimes referred to as the maximum ellipse of the circumscribing triangle, because it has the greatest area among all ellipses contained in the triangle. This result is apparently due to Jakob Steiner (see Problem 98 in [**2**]). Weisstein [**10**] refers to it to as the *Steiner Inellipse* of the circumscribing triangle, and describes several of its interesting properties.

Finally, the uniqueness of the inscribed ellipse is important in the statement of Marden's Theorem if we wish to view it as a recipe for constructing the roots of p' : start with the triangle, inscribe the ellipse, find the foci. This conceptualization would not make sense without uniqueness of the ellipse. Nevertheless, I do not regard the uniqueness of the inscribed ellipse as part of the content of the theorem, considering it rather as background knowledge. This is true to the spirit of both Bôcher and Marden, who treat the existence and uniqueness of such an ellipse as common knowledge. Bôcher even refers to it as the *maximum ellipse* of the triangle. Still, it should be recognized that a completely self-contained presentation of Marden's Theorem must include some discussion of the ellipse. Both existence and uniqueness are addressed in [**3**].

PROOFS OF MARDEN AND BÔCHER. For the sake of clarity, I will describe the logical gap in Marden's argument in the context of the theorem as I have stated it above. This corresponds to the case of the more general result where $m_j = 1$ for $j = 1, 2, 3$. Accordingly, let the polynomial $p(z)$, it's roots z_1, z_2, z_3, and the triangle T be as in my statement of Marden's Theorem. Marden shows that an ellipse with foci at the roots of p' and which passes through the midpoint of one side of T is actually tangent to the side there. By symmetry, this argument applies to any side of the triangle. That is as far as Marden's proof goes. What is missing is a verification that there is one ellipse that is tangent at all three midpoints. The appeal to symmetry establishes the existence of three ellipses, each with foci at the roots of p', and each tangent to one side of the triangle at its midpoint. To complete the proof, it must be shown that these three ellipses are actually one and the same.

In Bôcher's proof there is no logical gap. He shows that under the circumstances just described, an ellipse with foci at the roots of p' and which is tangent at the midpoint of one side of the triangle must also be tangent to the other two sides of the triangle. Invoking an earlier result showing that this ellipse must also be centered at the centroid of the triangle, he states: "Since only one conic can be drawn with a given point as centre and tangent to three given lines, this ellipse must be the maximum ellipse" (that is, the ellipse mentioned in Marden's Theorem). Alas, here I discover a gap in my understanding, even if there is no gap in the logic. As intuitively appealing as the statement is, I am not familiar with this uniqueness property, although I suspect it would be obvious to mathematicians of Bôcher's era, who were much more familiar with properties of curves than we are today. No doubt, with a little effort it would be possible to fill in this detail that Bôcher brushes off with a mere mention. On the other hand, it is appealing to combine elements of the proofs of Bôcher and Marden, because they are so similar. And as that leads to a complete and elementary proof, that is the route we shall follow.

In overview, here is the organization of the hybrid proof. First, following Marden, we will see that if an ellipse has foci at the roots of p' and if that ellipse passes through the midpoint of one side of the triangle T, then it is actually tangent to that side at the midpoint. This permits us to construct three ellipses, each of which is tangent to one side of the triangle at its midpoint. Now consider one of these ellipses. Following Bôcher, we will see that it is actually tangent to all three sides of the triangle. It remains only to show that this one ellipse coincides with the other two ellipses. But suppose this were not so. Then at one of the other two sides the point of tangency is not the midpoint. For that side there must therefore be a distinct ellipse that *is* tangent at the midpoint. But that implies that we have two distinct ellipses sharing the same foci, and tangent at distinct points of one line. Here I join Bôcher in saying that this is obviously impossible.

The complete presentation of the foregoing outline will center on proving two lemmas, one for Marden's part of the proof, and one for Bôcher's. In addition, we need one lemma about ellipses, and other preliminary observations. We turn to those next.

PRELIMINARIES. One important point to note is this: with no loss of generality, we may translate, rotate, and scale our triangle in any convenient manner. Furthermore, observe that these transformations can be imposed by a linear function $M : \mathbb{C} \to \mathbb{C}$. Let $M(z) = \alpha z + \beta$, where $\alpha \neq 0$ and β are fixed complex numbers, and express α in polar form $re^{i\theta}$. Then transforming z to $M(z)$ has the following geometric interpretation: scale z by r, rotate about the origin through angle θ, and translate by β. Thus any combination of scaling, rotating, and translating can be realized through the application of a linear function M.

We can now justify the claim that imposing these transformations represents no loss of generality in our proofs. That requires showing that Marden's theorem holds for a triple $\{z_1, z_2, z_3\}$ if and only if it also holds for the transformed triple $\{M(z_1), M(z_2), M(z_3)\}$. Actually, the invertibility of linear functions implies that we need only show one direction of the double implication. So, if we know that Marden's Theorem holds for $\{z_1, z_2, z_3\}$, let us see that it must also hold for $\{M(z_1), M(z_2), M(z_3)\}$.

In Figure 1, the triangle with vertices at the z_j is shown, with the inscribed ellipse tangent at the midpoints of the sides. The foci of the ellipse, which must be the roots of p', are also shown. Looking at M geometrically, we observe that scaling, rotating, and translating this figure preserve all of the ingredients. That is, the image of the triangle is a similar triangle, the image of the ellipse is still an inscribed ellipse tangent at the midpoints of the sides, and the images of the foci of the original ellipse are the foci of the transformed ellipse. For the transformed configuration, the new polynomial will be $p_M(z) = (z - M(z_1))(z - M(z_2))(z - M(z_3))$. The derivative is given by $p_M'(z)$. Our goal is to show that M carries the roots of p' to roots of p_M'.

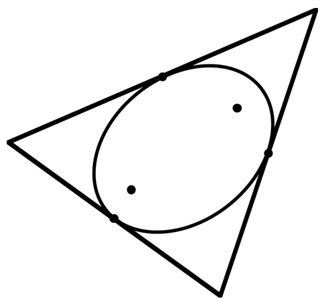

Figure 1. Geometric configuration for Marden's Theorem.

To that end, let us substitute $M(z)$ for z in the definition of p_M. That gives us

$$p_M(M(z)) = (M(z) - M(z_1))(M(z) - M(z_2))(M(z) - M(z_3)). \tag{1}$$

Next, notice that $M(z) - M(z_j) = \alpha(z - z_j)$. Therefore, equation (1) simplifies to

$$p_M(M(z)) = \alpha^3 p(z).$$

Now differentiate both sides of this equation, bearing in mind that $M'(z) = \alpha$. We obtain

$$\alpha p'_M(M(z)) = \alpha^3 p'(z)$$

and hence

$$p'_M(M(z)) = \alpha^2 p'(z).$$

This shows that if z is a root of p', then $M(z)$ is a root of p'_M. That is what we wished to show.

It is often helpful to recognize that the transformation M is invertible, allowing us to both do and undo the geometric transformations of translation, rotation, and dilation. For example, it was observed earlier that the special case of a double root of p' corresponds to an inscribed circle and an equilateral triangle. Since an invertible transformation is one-to-one, the preceding result now implies that p' has a repeated root if and only if p'_M does.

A second preliminary notion relates the coefficients and roots of a quadratic polynomial. As is well known, if the quadratic $z^2 + bz + c$ has roots z_1 and z_2, then the coefficients are given by $b = -(z_1 + z_2)$ and $c = z_1 z_2$. This will be needed in the proofs to follow.

The remaining preliminary topic concerns ellipses. Let us recall the familiar optical property of ellipses. Geometrically, this says that at any point of an ellipse, the tangent line makes equal acute angles with the lines to the two foci. We will need a less familiar, though equivalent, variant of this fact, which is stated in the following lemma.

Lemma 1. *Consider an ellipse with foci F_1 and F_2, and a point A outside the ellipse. There are two lines through A that are tangent to the ellipse. Let G_1 and G_2 be the points of tangency of these lines with the ellipse. Then $\angle F_1AG_1 = \angle F_2AG_2$.*

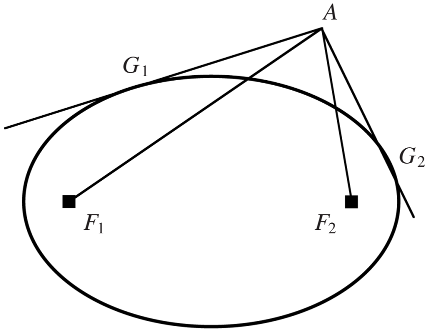

Figure 2. Ellipse, external point, and two lines.

Proof. The configuration described in the lemma is depicted in Figure 2, with one of two possible assignments of the labels G_1 and G_2. Observe that if the conclusion of the lemma holds for the labeling shown in the figure, it will also hold for the alternative labeling, and conversely. Therefore, there is no loss in generality to assign the labels as illustrated.

Reflect F_1 through AG_1 to define H_1, and let K_1 be the intersection of AG_1 with F_1H_1 (Figure 3). Then $\triangle AK_1F_1$ and $\triangle AK_1H_1$ are congruent right triangles, so $AF_1 = AH_1$ and $\angle F_1AK_1 = \angle H_1AK_1$. The analogous construction may be performed at G_2, leading to the analogous congruences. Now our object is to see that $\angle F_1AG_1 = \angle F_2AG_2$. Using the congruences just noted, it suffices to show $\angle F_1AH_1 = \angle F_2AH_2$.

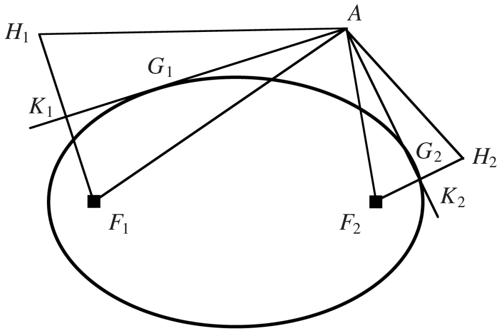

Figure 3. Isosceles triangles with apex at A.

Next draw the lines F_1G_1 and F_2G_1, extending the latter to H_1 (Figure 4). The collinearity of F_2, G_1, and H_1 is a consequence of the optical property of ellipses mentioned above. It shows $\angle F_1G_1K_1 = \angle F_2G_1A$. We also observe that $\angle F_1G_1K_1 = \angle H_1G_1K_1$ since AK_1 is the perpendicular bisector of F_1H_1. Thus $\angle F_2G_1A = \angle H_1G_1K_1$, showing that F_2, G_1, and H_1 are collinear. As before, an analogous construction occurs at G_2.

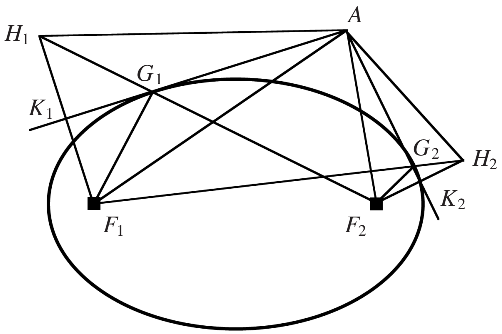

Figure 4. Using the optical property.

Now we will show that $\triangle AH_1F_2$ is congruent to $\triangle AF_1H_2$, by showing that their corresponding sides are equal. We have already observed $AH_1 = AF_1$ and $AF_2 = AH_2$. For the remaining two sides, we have $H_1F_2 = F_1G_1 + G_1F_2 = F_1G_2 + G_2F_2 = F_1H_2$, where we have used the fact that the sum of the distances from the foci to any point of the ellipse is constant.

Finally, the congruence of $\triangle AH_1F_2$ and $\triangle AF_1H_2$ implies that $\angle H_1AF_2 = \angle F_1AH_2$. Since these angles have $\angle F_1AF_2$ in common, we conclude that $\angle H_1AF_1 = \angle F_2AH_2$. As we observed earlier, this suffices to show that $\angle G_1AF_1 = F_2AG_2$, as desired. ■

Lemma 1 comes down to us from ancient Greek geometry. According to Milne and Davis [**7**], it appears as Proposition 46 in Book III of the *Conics* of Apollonius. Although it appears to be an extension of the optical property of ellipses, the two are actually equivalent. As shown above, the result for A outside the ellipse follows from the optical property. But if we assume the conclusion of the lemma, we can obtain the optical property as the limiting case with A approaching the ellipse.

Although this fact about ellipses is not part of the standard undergraduate curriculum, perhaps it should be. Its proof uses well known geometrical techniques, and it nicely extends the optical property that is usually at least mentioned in a unit on ellipses. It is also particularly interesting in the context of an ellipse inscribed in a triangle, where the role of A is played by any vertex, and the tangent lines are sides of the triangle. In this case, the lemma shows that the foci of the ellipse are *isogonal conjugates* (see [**9**]). As will become obvious in the sequel, the lemma is also an indispensible tool in our proof of Marden's Theorem.

THE MAIN EVENT. At last we are ready to prove Marden's Theorem. Following the outline discussed earlier, we will consider two lemmas, one based on the proof of Marden, and the other on the proof of Bôcher. We begin with Marden's contribution.

Lemma 2. *Let the polynomial $p(z)$, it's roots z_1, z_2, z_3, and the triangle T be as in the statement of Marden's Theorem. Then the ellipse with foci at the roots of p' and passing through the midpoint of one side of the triangle T is actually tangent to that side of T.*

Proof. As noted in the preceding section, with no loss of generality, we can rotate, translate, and scale the triangle in any way we choose. Accordingly, let us arrange things so that one side of the triangle lies along the x-axis centered at the origin, and has length 2, while the opposite vertex sits in the upper half-plane. Thus, the vertices of the triangle (and the roots of p) are at 1, -1, and $w = a + bi$ where $b > 0$. We will look at the ellipse passing through zero, which is the midpoint of the side that lies on the x-axis. In order to show that the ellipse is tangent to this side, we will show that the lines from the origin to each focus make equal angles with the x-axis.

Now we know the roots of p (which we may assume to be monic), so we have

$$p(z) = (z-1)(z+1)(z-w) = z^3 - wz^2 - z + w.$$

Differentiating, we find

$$p'(z) = 3z^2 - 2wz - 1 = 3\left(z^2 - \frac{2w}{3}z - \frac{1}{3}\right).$$

Thus, if the roots of p' are $z_4 = r_4 e^{i\theta_4}$ and $z_5 = r_5 e^{i\theta_5}$ (with $0 \le \theta_4, \theta_5 < 2\pi$), we conclude that $z_4 + z_5 = 2w/3$ and $z_4 z_5 = -\frac{1}{3}$. The first of these shows that at least one of the roots of p' must be in the upper half-plane, and the second then shows that $\theta_4 + \theta_5 = \pi$. This in turn also tells us that both roots must be in the upper half-plane. Moreover, considering these roots as vectors drawn from the origin, the angles the vectors make with the positive x-axis are supplementary. Thus, either both roots are on the y-axis, or one root makes an acute angle with the positive x-axis and the other roots makes an equal angle with the negative x-axis. In either case, this shows that the lines from the foci of our ellipse to 0 make equal angles with the x-axis, which is therefore a tangent line of the ellipse. ∎

On to Bôcher's contribution.

Lemma 3. *Let the polynomial $p(z)$, it's roots z_1, z_2, z_3, and the triangle T be as in the statement of Marden's Theorem. Consider the ellipse with foci at the roots of p' and which is tangent to one side of T at its midpoint. Then this same ellipse is tangent to the other two sides of T.*

Proof. As before, we are free to position the triangle any way we wish. The ellipse E is tangent to one side of the triangle, and we again place this side along the x-axis. However, this time we put one vertex at the origin, and the other at 1. The remaining vertex will again be placed at $w = a + bi$ where $b > 0$. This configuration will permit us to see that the ellipse is tangent to side $0w$.

With roots at 0, 1, and w, $p(z)$ can be taken as $z(z-1)(z-w)$. Multiplying out these factors gives

$$p(z) = z^3 - (1+w)z^2 + wz$$

and differentiation leads to

$$p'(z) = 3z^2 - 2(1+w)z + w.$$

By a similar argument as before, observe that $z_4 + z_5 = (2/3)(1+w)$. This shows that at least one of the roots of p' must be in the upper half-plane. But here we also know that these roots are the foci of an ellipse tangent to the x-axis. Therefore, both are in the upper half plane, allowing us to express the roots of p' as $z_4 = r_4 e^{i\theta_4}$ and $z_5 = r_5 e^{i\theta_5}$, where $0 < \theta_4 \le \theta_5 < \pi$.

Turning next to the constant term of p', we have $z_4 z_5 = w/3$. This shows that $\theta_4 + \theta_5$ is equal to the angle between the positive x-axis and Ow. Consequently, the angle between Oz_5 and Ow equals θ_4. See Figure 5.

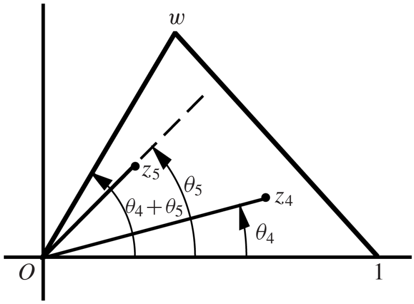

Figure 5. Configuration for Bôcher's result.

Now we apply Lemma 1, with the origin in the role of the external point A (Figure 6). How do we know the origin *is* external? This is apparent since E is tangent to the x-axis at $x = 1/2$. Indeed, the x-axis is one of the two tangent lines to E from the origin; let the other tangent line be L. Then by Lemma 1, the angle β between Oz_5 and L equals the angle between the x-axis and Oz_4, which in turn equals θ_4. But that is the same as the angle between Oz_5 and Ow. This shows that L is coincident with Ow, and thus, Ow is tangent to E.

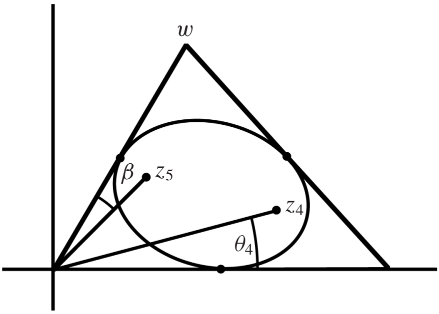

Figure 6. The ellipse is tangent to $0w$ because $\beta = \theta_4$.

It remains to show that the side from 1 to w is also tangent to the ellipse. This can be established by essentially the same proof, but with the triangle translated horizontally so that it has vertices at -1 and 0, rather than at 0 and 1. Carrying out this plan completes the proof of the lemma. ■

And now, finally, we arrive at the proof of Marden's Theorem.

Proof of Marden's Theorem. As usual, we assume that polynomial p, its roots z_j, and triangle T are as in the statement of the theorem. Using the roots of p' as foci, draw an ellipse E that passes through the midpoint of one side of the triangle. By Lemma 2, E is actually tangent to that side of T. By Lemma 3, E is also tangent to the other two sides of T. Now we claim that the points of tangency with these other two sides must be the midpoints. If not, repeat the construction above with a new side, producing an ellipse E'. Since E and E' have the same foci, and are both tangent to the same line (in fact to three lines), they must actually coincide. But this would show that E and E' both contact the new side in the same point, namely, the midpoint. By symmetry, the same conclusion holds for the remaining side of the triangle. Thus, the original ellipse E is tangent to all three sides at their midpoints. That completes the proof of Marden's Theorem. ■

REFERENCES

1. M. Bôcher, Some propositions concerning the geometric representation of imaginaries, *Ann. of Math.* **7** (1892) 70–76.
2. H. Dörrie, *100 Great Problems of Elementary Mathematics, Their History and Solution* (trans. D. Antin), Dover, New York, 1965.
3. D. Kalman, The most marvelous theorem in mathematics, *Journal of Online Mathematics and its Applications*, available at `http://www.JOMA.org` (for journal), `http://mathdl.maa.org/mathDL/4/?pa=content&sa=viewDocument&nodeID=1663` (direct to article).
4. B. Z. Linfield, On certain polar curves with their applications to the location of roots of derivatives of a rational function, *Trans. Amer. Math. Soc.* **27** (1920) 17–21.
5. M. Marden, A note on the zeros of the sections of a partial fraction, *Bull. Amer. Math. Soc.* **51** (1945) 935–940.
6. ———, *Geometry of Polynomials,* Math. Surveys no. 3, American Mathematical Society, Providence, RI, 1966.
7. J. J. Milne and R. F. Davis, *Geometrical Conics,* MacMillan, New York, 1894.
8. J. Siebeck, Ueber eine neue analytische Behandlungweise der Brennpunkte, *J. Reine Angew. Math.* **64** (1864) 175–182.
9. E. Weisstein, Isogonal Conjugate—From MathWorld, A Wolfram Web Resource, `http://mathworld.wolfram.com/IsogonalConjugate.html`.
10. ———, Steiner Inellipse—From MathWorld, A Wolfram Web Resource, `http://mathworld.wolfram.com/SteinerInellipse.html`.

DAN KALMAN received his Ph.D. from the University of Wisconsin in 1980. Before joining the mathematics faculty at American University in 1993, he worked for eight years in the aerospace industry in Southern California. Kalman is a past associate executive director of the MAA, author of a book published by the MAA, and frequent contributor to MAA journals. He delights in puns and word play of all kinds, and is an avid fan of Douglas Adams, J. R. R. Tolkein, and Gilbert and Sullivan.
Department of Mathematics and Statistics, American University, 4400 Massachusetts Avenue NW, Washington, DC 20016-8050
kalman@american.edu

A Lemma on Divisibility

Proofs of the unique factorisation theorem for integers tend to follow one of two routes. One method is to show from the Euclidean algorithm that

$$\text{for all primes } p,\ p \mid ab \text{ implies either } p \mid a \text{ or } p \mid b,$$

from which the theorem follows quickly. The other method uses an ingenious induction over the integers. Both methods can be found, for instance, in the books of Davenport (*The Higher Arithmetic: An Introduction to the Theory of Numbers*) and Hardy and Wright (*An Introduction to the Theory of Numbers*), and in very many other places. Thus it is interesting to give an independent direct proof of the statement displayed above, using an induction over the primes.

We say that the prime p has property P if for all integers a and b, $p \mid ab$ implies either $p \mid a$ or $p \mid b$. Clearly 2 has property P, since if ab is even then one of a and b must be even. This sets up the induction. Now supppose that for some prime p, all primes strictly less than p have property P, and suppose that $p \mid ab$. We can write $a = mp + c$ and $b = mp + d$, where $0 \leq c, d < p$ and $p \mid cd$. If either $c = 0$ or $d = 0$, then p divides a or b as required. If not, then both c and d are at least 1 and so can be factored into primes strictly less than p, say $c = \prod p_i^{a_i}$, $d = \prod p_i^{b_i}$ (*existence* of factorisation has of course been proved earlier). Now since $p \mid cd$ there is some n such that $np = \prod p_i^{a_i+b_i}$, where all p_i are less than p and so have property P. But since p is also prime, it cannot be divisible by p_i, and so each p_i must divide n. Thus we may cancel all the primes p_i successively, ending up with some integer n' for which $n'p = 1$, which is clearly impossible.

—Submitted by Peter Walker, American University of Sharjah,
Sharjah, United Arab Emirates
peterw@aus.edu