The Dynamics of Newton's Method for Cubic Polynomials

James A. Walsh

James Walsh, upon receiving his B.S. from the University of Connecticut, taught for two years in the Peace Corps in Togo, West Africa. He also taught high school mathematics and received his M.A. in education from Fairfield University. After receiving his Ph.D. in mathematics from Boston University in 1991 he came to Oberlin College, where he is an assistant professor. His research interests are in dynamical systems. He is also interested in incorporating aspects of dynamical systems theory into the undergraduate curriculum. In his spare time he and his wife Debbi enjoy watching their son Zach crawl around the floor.

The aim of this article is to investigate the behavior of Newton's method for cubic polynomials with real coefficients—a seemingly simple task! Yet we hope the reader will find the results not only of interest but somewhat surprising as well. We begin with the basic ideas and terminology of *iteration*.

Given $F\colon \mathbf{R} \to \mathbf{R}$ and $x_0 \in \mathbf{R}$, the sequence

$$x_0, F(x_0), F(F(x_0)) = F^2(x_0), F(F(F(x_0))) = F^3(x_0), \ldots, F^n(x_0), \ldots$$

is called the *orbit* of x_0. The term $F^n(x_0)$ represents n-fold composition of F and is the nth iterate of x_0. The basic question one poses in studying iteration is "What happens to orbits over time (as $n \to \infty$)?"

We first focus on the simplest orbits. If $F(x_0) = x_0$, so that x_0 is a *fixed point* of F, the orbit of x_0 is $x_0, x_0, x_0, \ldots$. More generally, if $F^n(x_0) = x_0$ for some $n \geq 0$, then x_0 lies on a *periodic orbit* (or *cycle*), an orbit that simply repeats over time. Periodic orbits also come equipped with certain dynamical properties; if $F^n(x_0) = x_0$ and $|(F^n)'(x_0)| < 1$, then orbits with initial points near x_0 converge to the orbit of x_0. We say that such an x_0 lies on an *attracting* periodic orbit. If x_0 is a period n point for F and $|(F^n)'(x_0)| > 1$, then nearby orbits begin by moving away from the orbit of x_0. In this case the orbit of x_0 is said to be a *repelling* periodic orbit [5].

If x_0 lies on an attracting periodic orbit, the *basin of attraction* of x_0 is the set of all points whose orbits converge to the orbit of x_0 as $n \to \infty$. We will soon see that basins of attraction can be rather complicated sets! The *immediate basin of attraction* $W(x_0)$ of a periodic point x_0 is the largest *interval* containing x_0 that lies entirely in the basin of attraction of x_0.

Given a differentiable function $f\colon \mathbf{R} \to \mathbf{R}$, Newton's method consists of iterating the function $N_f(x) = x - f(x)/f'(x)$. Evidently the roots of f are fixed points of N_f, and we would like to determine the possible behaviors of orbits when N_f is iterated. The familiar geometry of Newton's method is shown in Figure 1.

If x_0 is a root of f of multiplicity one, i.e., $f(x_0) = x_0$ but $f'(x_0) \neq 0$, then $N_f'(x_0) = f(x_0)f''(x_0)/(f'(x_0))^2 = 0$, so x_0 is an attracting fixed point for N_f. In fact, if x_0 is a root of f of multiplicity greater than one, it is still true that x_0 is an attracting fixed point of N_f [5, p. 167]. The attractive nature of these fixed points is

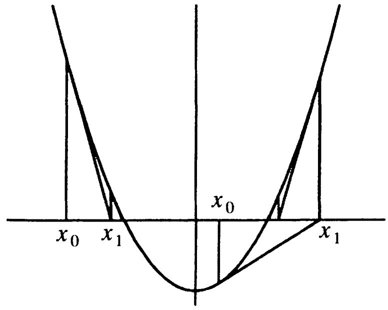


Figure 1
The geometry of Newton's method.

what makes Newton's method "work"; if we begin the iteration of N_f with an initial point in the basin of attraction of a root x_0 of f, then the orbit of this point will converge to x_0. Thus finding the basin of attraction of each root, or at least the immediate basin of attraction—the largest interval containing the root inside which Newton's method will always converge to that root—is an important part of the study of Newton's method. We would also like to determine the set of initial values for which Newton's method fails, i.e., the points that are not in the basin of attraction of any root of f. That is, we want to find the set

$$E = \left\{x \in \mathbf{R}\text{: the sequence } N_f^n(x) \text{ does not converge to a root of } f \text{ as } n \to \infty\right\}.$$

To simplify matters we will consider functions only as complicated as cubic polynomials. As a warm-up we begin with the simplest of all functions, namely $f(x) = ax + b$ where $a \neq 0$. Orbits for N_f are extremely well behaved—given any $x_0 \in \mathbf{R}$, $N_f(x_0)$ is the solution of $f(x) = 0$ (of course, there is a simpler way to solve $ax + b = 0$!). We note that $E = \emptyset$.

Suppose f is a quadratic polynomial with distinct real zeros. Denote these zeros by l and r, $l < r$, and let e be the critical point of f. Again orbits are very well behaved for N_f. The basin of attraction of r coincides with the immediate basin of attraction of r, which is the interval $(e, +\infty)$. To see this, note that $x_0 > r$ implies $N_f^n(x_0) < N_f^{n-1}(x_0)$ and $N_f^{n-1}(x_0) > r$ for all $n \geq 1$. Hence the orbit of x_0 converges to some point p as $n \to \infty$, and by the continuity of N_f, $N_f(p) = p$. Hence $p = r$. If $e < x_0 < r$, $N_f(x_0) > r$, so $W(r) = (e, +\infty)$; see Figure 1. Likewise the basin of attraction of l is the interval $(-\infty, e)$. In this setting $E = \{e\}$, so that if $x_0 \neq e$ the orbit of x_0 converges to either l or r.

We now consider cubic polynomials, and again we simplify matters before proceeding. Note that if $g(x) = kf(x)$ for some constant k, then $N_f(x) = N_g(x)$, so we consider only monic cubics. Also, if $g(x) = f(A(x))$ where $A(x) = ax + b$, then $AN_gA^{-1}(x) = N_f(x)$. That is, N_g and N_f are conjugate by an affine transformation and thus exhibit the same dynamics. If x_0 lies on a periodic orbit for N_g, then $A(x_0)$ lies on a periodic orbit for N_f. Moreover, if x_0 lies on an attracting (resp. repelling) cycle, $A(x_0)$ lies on an attracting (resp. repelling) cycle. Hence an affine change of variables does not alter the dynamics of Newton's method, so for our family of cubics we will fix one zero at $x = -2$ and take $\pm\sqrt{-c}$ as the other two zeros, where $c \in \mathbf{R}$. Our family of cubics is then $f_c(x) = (x+2)(x^2+c)$, with associated Newton's method function

$$N_c(x) = x - \frac{f_c(x)}{f_c'(x)} = \frac{2x^3 + 2x^2 - 2c}{3x^2 + 4x + c},$$

a one-parameter family of rational maps. For future reference, we note that in general $N_f'(x) = f(x)f''(x)/(f'(x))^2$; thus the critical points of N_c are the three roots of f_c together with $x = -\frac{2}{3}$, the root of f_c''.

We first consider f_c for $c < 0$, so that $f_c(x) = 0$ has three distinct, real solutions. Once again the dynamics of N_c, while more interesting than in the quadratic case, are still rather tame. We present a brief analysis for the case $c = -1$, setting $f(x) = f_{-1}(x)$ and $N(x) = N_{-1}(x)$; see also **[7]**, **[10]**.

As in the quadratic case, $W(1) = (e_2, +\infty)$ and $W(-2) = (-\infty, e_1)$, where $e_1 < e_2$ are the critical points of f. Since $x = -1$ is an attracting fixed point for N and N is continuous on (e_1, e_2), $W(-1)$ is an open interval (a, b) for some a, b with $e_1 < a < b < e_2$. Moreover, by the definition of the immediate basin of attraction, $N((a,b)) = (a,b)$, and hence there are four possibilities as to the images of a and b under N. Either $N(a) = a$ and $N(b) = b$, $N(a) = b$ and $N(b) = a$, $N(a) = a$ and $N(b) = a$, or $N(a) = b$ and $N(b) = b$. Three of these possibilities imply that either a or b (or both) is fixed by N, a contradiction. Thus $N(a) = b$ and $N(b) = a$, so that a and b form a two-cycle for N. As we will see shortly, this is a repelling two-cycle, which of course is good news for Newton's method! A graph of $y = N^2(x) - x$ shows that the equation $N^2(x) = x$ has five real solutions, three of which are the fixed points $x = -2$, $x = -1$, and $x = 1$. The remaining two solutions must be a and b. We note also that $a < -\frac{2}{3} < b$.

A computation yields $N'(x) < 0$ on $(e_1, -1) \cup (-\frac{2}{3}, e_2)$. Thus N maps (e_1, a) in a one-to-one fashion onto $(b, +\infty)$. Choose $e_3 \in (e_1, a)$ such that $N(e_3) = e_2$. Likewise, let $e_4 \in (b, e_2)$ be the unique point satisfying $N(e_4) = e_1$ (see Figure 2).

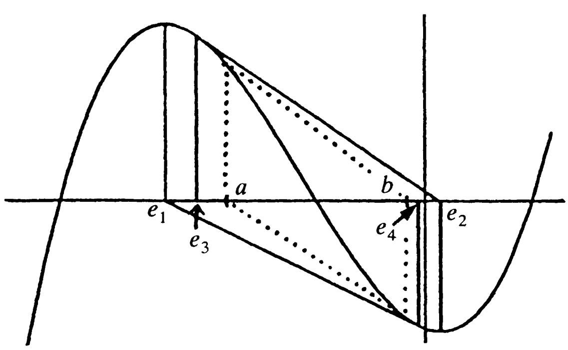


Figure 2
The two-cycle $\{a, b\}$ and the beginnings of the sequence $\{e_k\}$.

Now note that $N((e_3, a)) = (b, e_2)$, so choose $e_5 \in (e_3, a)$ such that $N(e_5) = e_4$. Similarly, let $e_6 \in (b, e_4)$ satisfy $N(e_6) = e_3$. Continuing in this fashion, we construct a recursively defined sequence $\{e_k\}_{k=1}^{\infty}$ with $N(e_{2n+1}) = e_{2n}$ and $N(e_{2n+2}) = e_{2n-1}$ for $n \geq 1$. Since $e_1 < e_3 < e_5 < \cdots < a$ and $e_2 > e_4 > e_6 > \cdots > b$, the subsequence $\{e_{2n-1}\}_{n=1}^{\infty}$ converges to a limit $\leq a$ and $\{e_{2n}\}_{n=1}^{\infty}$ converges to a limit $\geq b$.

We claim that $e_{2n-1} \to a$ and $e_{2n} \to b$ as $n \to \infty$. Let $g = N^{-1} \circ N^{-1}$: $(b, e_2) \to (b, e_2)$. For all $n \geq 1$, $g(e_{2n}) = e_{2n+4}$, and so by the continuity of g, $e_{2n} \to p$ as

$n \to \infty$ implies $g(p) = p$. That is, p is a period-two point for N^{-1}, and hence also for N. We have seen that the period-two points for N are a and b, implying $p = b$. Likewise, $e_{2n-1} \to a$ as $n \to \infty$.

The reader may well be wondering why we choose to focus on the images under N^{-1} of the critical points of f. The answer is that these inverse images form the boundaries of the connected components of the basins of attraction of -2 and 1 as follows. $N((e_1, e_3)) = (e_2, +\infty)$ and hence (e_1, e_3) is contained in the basin of attraction of $x = 1$. $N((e_4, e_2)) = (-\infty, e_1) = W(-2)$, so the orbit of any $x \in (e_4, e_2)$ converges to $x = -2$ as $n \to \infty$. Likewise, $N((e_6, e_4)) = (e_1, e_3)$, and so (e_6, e_4) is a subset of the basin of attraction of $x = 1$. Continuing in this fashion and focusing on the interval (e_1, a), we see that (e_{4n-3}, e_{4n-1}) is contained in the basin of attraction of $x = 1$, while (e_{4n-1}, e_{4n+1}) is a subset of the basin of attraction of $x = -2$ for all $n \geq 1$. On the interval (e_1, a) we have an infinite sequence of alternating components of the basins of attraction of $x = 1$ and $x = -2$, respectively. The lengths of these intervals decrease to 0, with the endpoints converging to a. A similarly complex pattern exists on the interval (b, e_2); see Figure 3. An example in [**11**] shows that the intervals forming each basin of attraction decrease approximately geometrically in length.

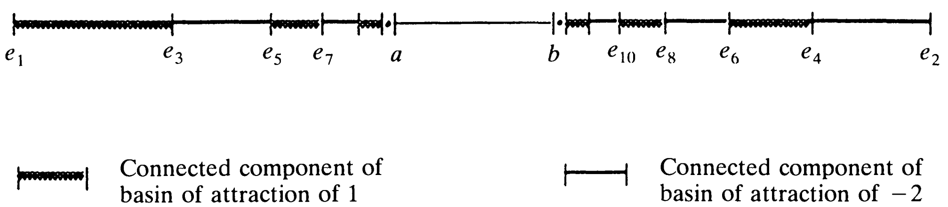


Figure 3

As mentioned above, the dynamics of iterating N_c for $c < 0$ are more complex than in the case of Newton's method for quadratic polynomials, yet remain tame. In particular, the set E just contains the sequence $\{e_k\}_{k=1}^{\infty}$ and a repelling two-cycle $\{a, b\}$, so that E is a countable set. Newton's method will still converge to a zero of $f_c(x) = 0$ on a set of initial points having full Lebesgue measure in $\mathbf{R}$. We note the contrast to the case where f is a polynomial of degree $d \geq 4$ with d real, distinct zeros. In this setting E contains a Cantor set [**1**], and N_f restricted to this Cantor set is chaotic [**8**].

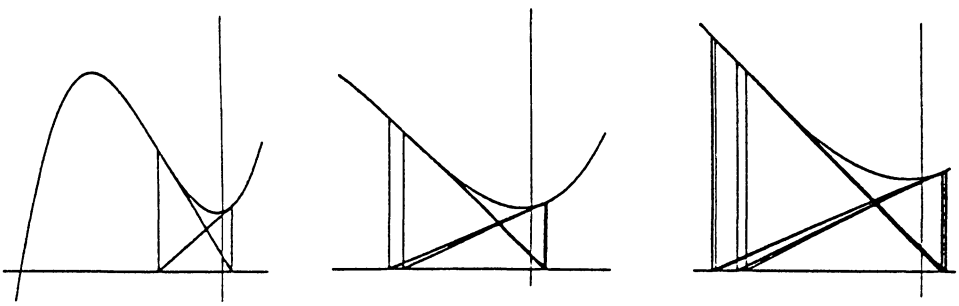
We now turn our attention to computer experiments carried out in the case $c > 0$, so that $f_c(x) = 0$ has one real and two complex solutions. We will make use of the following theorem of P. Fatou [**6**].

Theorem. *If $R(z)$, a rational function of a complex variable z, has an attracting periodic cycle, then the orbit of at least one critical point will converge to it.*

Remark. In the case of our cubic polynomial f with $c < 0$, we have seen that the rational function $N_c(x)$ has a two-cycle $\{a, b\}$, and that the critical points of N_c are the three real roots of f (fixed points of N_c) and the point $x = -\frac{2}{3}$. Since $a < -\frac{2}{3} < b$, and (a, b) is the immediate basin of attraction of the middle root, the orbit of $-\frac{2}{3}$ converges to this root. Hence the theorem implies that the two-cycle

$\{a, b\}$ cannot be attracting. Our earlier analysis showed that it is in fact repelling: All orbits that start near a or b converge to one of the three roots of f, unless they start at one of the points e_k, in which case the orbit moves away until it reaches either e_1 or e_2, where N_c is undefined.

Since $x = -\frac{2}{3}$ is the only critical point of N_c that is not fixed under N_c, Fatou's theorem implies that if N_c has an attracting periodic orbit then the orbit of $x = -\frac{2}{3}$ will converge to it. Thus we will examine the orbit of $x = -\frac{2}{3}$ under N_c as c varies through positive values, hoping to find values of c for which this orbit is attracted to a periodic orbit. For example, as c decreases from 0.2, plotting only the long-term behavior of the orbit we find the graphs shown in Figure 4: N_c passes from having an attracting two-cycle to an attracting four-cycle, then an attracting eight-cycle. (The periods may be difficult to discern in the graphs, but lists of numerical values of $N_c^n(-\frac{2}{3})$ make it clear that they are 2, 4, and 8.)



$c = 0.196$ $c = 0.190$ $c = 0.188$

Figure 4
Periodic orbits that attract the orbit of $x = -\frac{2}{3}$.

Evidently N_c is undergoing a sequence of *period-doubling bifurcations* [**5**]. This naturally leads us to consider a *bifurcation diagram*, where the values of c are placed on the vertical axis and, for a fixed c, iterates 100 through 200 of $x = -\frac{2}{3}$ are plotted horizontally; see Figure 5.

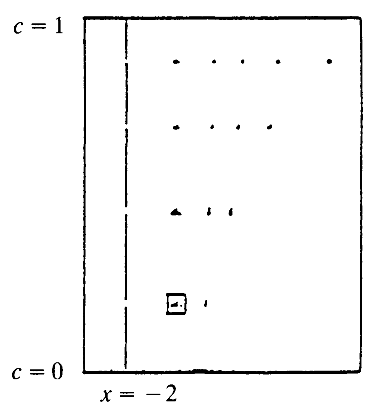


Figure 5
The bifurcation plot for N_c.

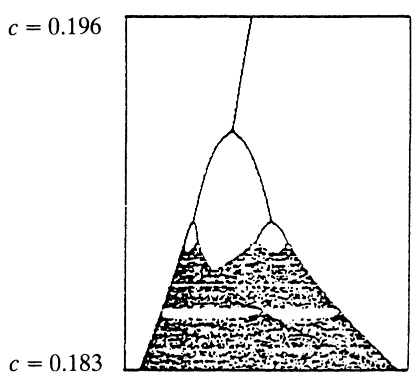


Figure 6
Boxed portion of Figure 5, enlarged.

We first notice that for most positive values of c the orbit of $x = -\frac{2}{3}$ converges to $x = -2$. The exception is the sequence of small gaps in the values of c where the orbit is doing "something else." Zooming in on the box in Figure 5 yields the striking (and most likely familiar) bifurcation plot in Figure 6, normally associated with quadratic maps [**5**] (for complex cubics a similar experiment yields copies of the Mandelbrot set in parameter space [**4**]; see also [**2**], [**3**]).

This suggests (since there are in fact two copies of this plot side by side in Figure 5) that N_c^2 is "quadratic like" on some interval. Figure 7 shows successive plots of N_c^2 for $c = 0.193$, 0.188, and 0.175. We note the similarity with [**5**, p. 90]. We also remark that there exist infinitely many of these copies of the "quadratic" bifurcation plot within the bifurcation plot for N_c (some of which can be discerned in Figure 5).

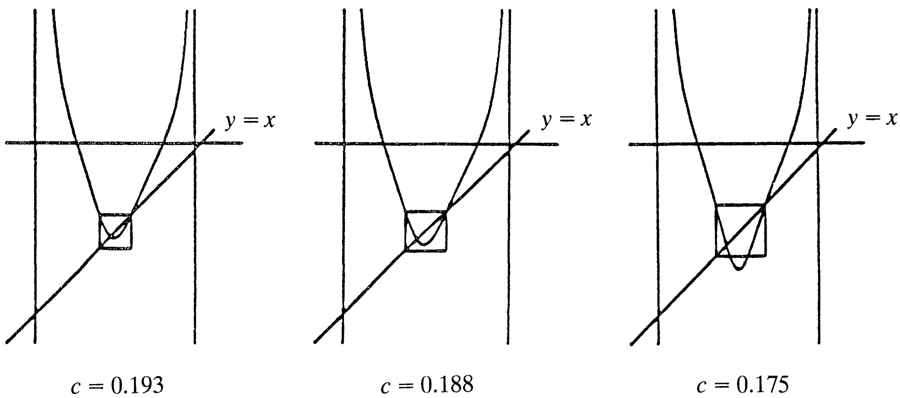


Figure 7
N_c^2 is locally "quadratic like."

As a final aid in understanding the dynamics of N_c, we use the following theorem [**8**], presented here in a less general fashion appropriate for our family f_c. The Sarkovskii ordering [**9**] of the natural numbers referred to in the theorem is

$$3 \rhd 5 \rhd 7 \rhd \cdots \rhd 2 \cdot 3 \rhd 2 \cdot 5 \rhd 2 \cdot 7 \rhd \cdots \rhd 2^2 \cdot 3 \rhd 2^2 \cdot 5 \rhd 2^2 \cdot 7 \rhd \cdots \rhd 2^2 \rhd 2 \rhd 1.$$

Theorem. *Let f be a polynomial and N_f the associated Newton's method map. If N_f has an orbit of period $n > 1$ then at least one of the following holds*:

(1) *For every integer m where $n < m$, N_f has an orbit of period m.*

(2) *For every integer m where $n \rhd m$, N_f has an orbit of period m.*

Let's see how our results for cubic polynomials fit into the general picture this theorem provides of the dynamics of Newton's method for polynomials. First, in the case $c < 0$, recall that we've shown that N_c has a period-two orbit. The theorem asserts that either for all $m > 2$ N_c has an orbit of period m (which we know is not the case) or else for all $m \lhd 2$ N_c has an orbit of period m. But 1 is the only integer satisfying the latter "inequality," so the theorem just asserts that N_c has at least one fixed point. And indeed this is true: N_c has three fixed points, namely the roots of f.

For $c > 0$, however, the orbit structure is much more complex. For $c = 0.46$, for example, we find a three-cycle (Figure 8). The above theorem then guarantees the existence of cycles of period m for all $m > 3$ or all m such that $3 \rhd m$, and every natural number satisfies this second "inequality." In either case, there exist

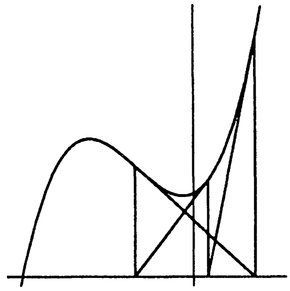


Figure 8
Period three implies *very* complicated dynamics.

infinitely many periodic orbits for N_c with different periods. Points on these periodic orbits of period $m > 1$ are members of the set E, so we see that the nature of this set is dramatically more complex for certain positive values of the parameter c. Further analysis shows that in this case there are uncountably many orbits that remain bounded yet do not converge to any periodic orbit, so that the set E of starting points for which Newton's method will fail is uncountable [**12**].

We hope we have shown that Newton's method, even for a family of real cubic polynomials, contains surprisingly rich dynamical behavior.

Author's note. This article was motivated by a software program I wrote while employed at Bolt, Beranek and Newman, Inc., in Cambridge, MA. The program allows the user to actively explore the dynamics of N_c for the above family of cubics f_c, including the bifurcation diagram. In the first semester calculus course I devote four classes to iteration, with the fourth class a computer lab based on this software. My students become quite engaged by Newton's method.

Acknowledgments. I thank Paul Blanchard for suggesting the family f_c, and Paul Horwitz and Wally Feurzeig at BBN.

References

1. B. Barna, Uber die divergenzpunkte des Newtonschen Verfahrens zur Bestimmung von Wurzeln algebraischer Gleichungen II, *Publicationes Mathematicae*, Debrecen, 4 (1956) 384–397.
2. P. Blanchard and D. Berkey, *Calculus*, 3rd ed., Saunders, Philadelphia, 1992, Appendix 4.
3. P. Blanchard and A. Chui, Complex dynamics: An informal discussion, in *Fractal Geometry and Analysis*, J. Belair and S. Dubuc, eds., Kluwer, Boston, 1991, 45–98.
4. J. Curry, L. Garnett, and D. Sullivan, On the iteration of a rational function: Computer experiments with Newton's method, *Communications in Mathematical Physics* 91 (1983) 267–277.
5. R. L. Devaney, *A First Course in Chaotic Dynamical Systems*, Addison-Wesley, Reading, MA, 1992.
6. P. Fatou, Sur les équations fonctionelles, *Bulletin Société Mathématique de France* 45 (1919) 161–271.
7. J. Hubbard and B. West, *Differential Equations: A Dynamical Systems Approach*, Springer-Verlag, New York, 1991, p. 230.
8. D. Saari and J. Urenko, Newton's method, circle maps, and chaotic motion, *American Mathematical Monthly* 91 (1984) 3–17.
9. A. N. Sarkovskii, Coexistence of cycles of a continuous map of a line into itself, *Ukrainskii Matematicheskii Zhurnal* 16 (1964) 61–71.
10. P. Straffin, Jr., Newton's method and fractal patterns, *UMAP Journal* 12.2 (1991) 143–164.
11. Philip Straffin, Newton's method and fractal patterns, in P. Straffin, ed., *Applications of Calculus* (vol. 3 of *Resources for Calculus Collection*, MAA Notes vol. 29), Mathematical Association of America, Washingon, DC, 1993, 68–84; pp. 72–74.
12. J. Yorke and T. Li, Period three implies chaos, *American Mathematical Monthly* 82 (1975) 985–992.