Questioning Answers

B. Mazur

[These are preliminary notes for the "Arnold Ross Lecture" to be given April 27 1996 at the University of Maryland]

All of us who are fascinated by mathematics have the faith (backed somewhat by experience) that math answers questions, solves problems. We work hard on some problem in mathematics because we are pursued by curiosity for the answer. Or, for the lesser reason that it was a problem that someone posed for us, perhaps as a challenge, perhaps as a test. We work hard, and then when we get "the answer" we might imagine that we can relax. I want to spend this hour turning this picture upside-down to suggest that much of the art of mathematics only begins once we have "the answer". If we can manage, at that point, to ask the right *questions* of the answer that we have gotten, we may be led to even more interesting things.

Let us start with a frivolous-sounding question about numbers and spend the hour investigating the realms of mathematics that the solution(s) to this question "invite" us to investigate[1].

Question: The number 210 is both the product of two consecutive integers, $210 = 14 \cdot 15$, and is also the product of three consecutive integers, $210 = 5 \cdot 6 \cdot 7$. How many other numbers have this property of being expressible as both the product of three consecutive integers and the product of two consecutive integers?

Before I begin dealing directly with this question, I want to chat a bit about its nature. Why did I choose it to talk about? What do I

[1] I will not try to prove things systematically, but I will call upon you to make a few calculations at various times

have up my sleeve? But as I do this, those of you who can "do background computing" in your heads (I can't !!) might try to find some of the other numbers that are the products of two and of three consecutive integers.

Our question, of course, can be rephrased as an algebraic problem. Thinking of our number N as the product of three consecutive integers, let X be the middle one of those three integers and we have

$$N = (X-1)\cdot X\cdot(X+1) = X^3 - X.$$

Thinking of N as the product of two consecutive integers, let Y be the smaller of those two integers, so we get

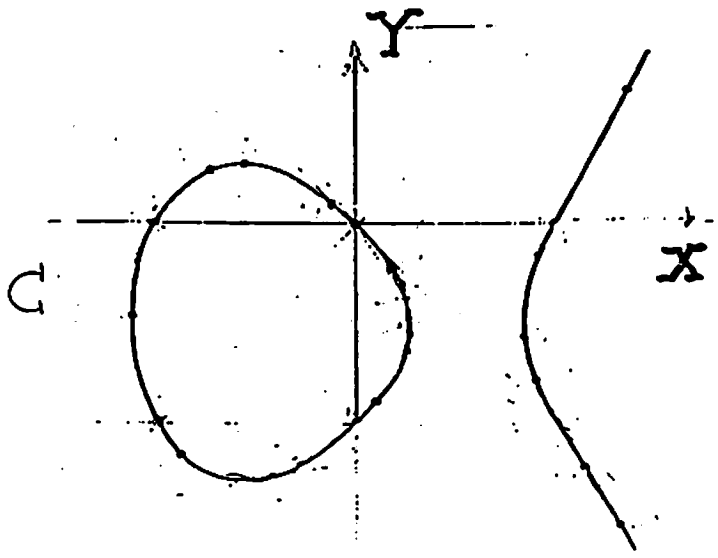
$$N = Y\cdot(Y+1) = Y^2+Y.$$

We have a solution "N" to our problem, then, every time we can find a pair of integers [X,Y] which have the property that

$$(*) \qquad Y^2+Y = X^3 - X.$$

We are faced, then, with an equation in two variables X and Y whose highest degree term is a cube, and we are looking for solutions to this equation in whole numbers. We can visualize the real solutions to this equation as a curve C in the (X,Y)-plane

---Graph of the equation $Y^2+Y = X^3 - X$ --



and we are looking for points on the curve C with integral (X,Y)-coordinates.

To "weigh" this problem confronting us, it might pay to compare it to a much simpler equation that is the "standard fare" of high school algebra: and is familiar to all of us:

The quadratic equation in one variable. Find the values of the variable X that "solves" the quadratic equation

$$a \cdot X^2 + b \cdot X + c = 0.$$

We all know the gambit here-- an idea which has come down to us from Babylonian times: if we want to find numbers X which solve this equation (and we might be interested in integer solutions X, or rational numbers X, or real numbers, or complex numbers) we "complete the square" by rewriting this equation as

$$a \cdot (X+b/2a)^2 + (c - b^2/4a) = 0,$$

and this rewritten equation *visibly* has (at most) 2 beautiful solutions given by the quadratic formula. Of course, if we are specifically interested in integer solutions, or rational solutions, we must check whether the answers given by our quadratic formula are integers or rational, etc. I said "of course" in the last sentence, but I should remind you that this issue of whether or not the answers "given by our quadratic formula are integers or rational, etc." was historically, at least, not such a humdrum affair. For example, the fact that $X^2 - 2$ has no rational solutions in X (that is, the fact that the square-root of 2 is irrational) was viewed as devastating by the Pythagorean mathematicians who initially made this discovery. The irrationality of the square-root of 2 was considered to be such a dark secret about the universe that when one of them revealed it to outsiders the story goes that he was murdered as a betrayer.

This may be so, or may not be so, but one thing is certain: for the

quadratic equation there are *at most* 2 solutions, and for any polynomial equation in one variable X of degree d, there are *at most* d solutions. One of the fascinations of the type of problem posed by the equation $Y^2+Y = X^3 - X$ that we are considering is that we don't even have any idea *how many solutions to expect* !

How many solutions have you found?

Let us start with some modest solutions to (*) that are so modest, you might have overlooked them:

$X = 0,\ Y = 0,$
$X = 0,\ Y = -1,$
$X = \pm 1,\ Y = 0,$
$X = \pm 1,\ Y = -1.$

All of these solutions give **N = 0**:

$0 = 0 \cdot 1 \cdot 2 = 0 \cdot 1$
$= 0 \cdot 1 \cdot 2 = (-1) \cdot 0.$

We will come back to the issue of exactly how modest or immodest these solutions are, later.

Did you also discover

$X = 2,\ Y = -3,$
$X = 2,\ Y = 2$?

These solutions give **N = 6**:

$6 = 1 \cdot 2 \cdot 3 = (-3) \cdot (-2)$
$1 \cdot 2 \cdot 3 = 2 \cdot 3.$

And then there are the solutions which were given in the statement of the problem itself:

X= 6, Y = -15,
X =6, Y = 14 ?

These solutions give N = 210:

$210 = 5 \cdot 6 \cdot 7 = (-15) \cdot (-14)$
$= 5 \cdot 6 \cdot 7 = 14 \cdot 15.$

Did you find any others? Suppose you had tried out all numbers under a million and had found no further solutions. Would you then be confident that there were no further ones? Now although "confidence" is a precious virtue that counts for a lot in mathematical work, my recommendation, in exactly this sort of calculation is that you should **NOT** be confident that you have gotten all the solutions. Let me illustrate why, by bringing up a slightly different problem to ours. **Problem:** Find the integer solutions to the equation

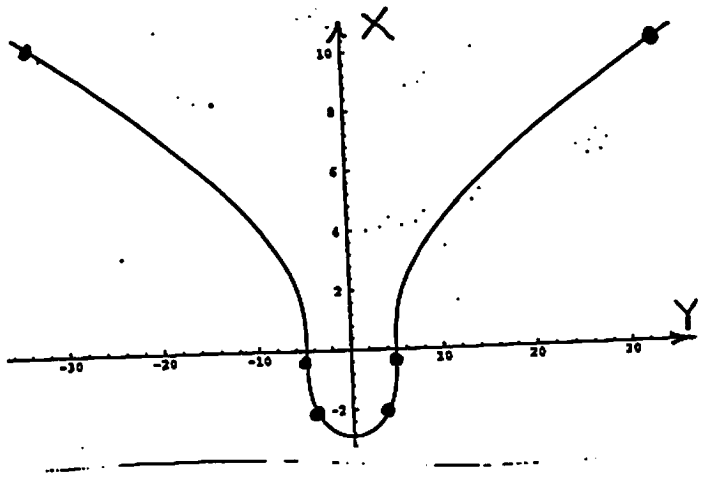
$$Y^2 = X^3 + 24;$$

that is, find the perfect squares ("Y^2") which are 24 more than a perfect cube ("X^3"). Now you will surely easily guess a few of the solutions. For example, X= -2, X=1, and X = 10 give solutions to this problem:

$$4^2 = (-2)^3 + 24$$
$$5^2 = 1^3 + 24$$
$$32^2 = 10^3 + 24.$$

But these are **NOT** all: there is one missing value of X that solves the equation and if, we wanted to plot that value of X on the graph below we would have to extend the "wingspan" of the graph (from the eight and a half inches that it takes up on the transparency) to a diameter of 20 miles!

---Graph of the equation $Y^2 = X^3 + 24$ ---



A "basic symmetry" in the equation $Y^2+Y = X^3 - X$: The first thing that jumps to the eye, given the solutions of $Y^2+Y = X^3 - X$ that we have already found is that these solutions come in pairs, each pair giving the same value for N. In fact, the entire curve C is brought to itself by the "symmetry"

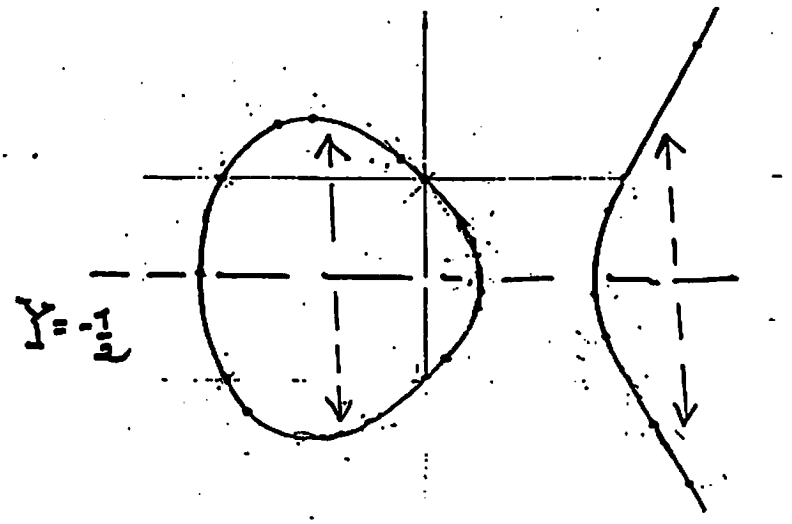
$$X \mapsto X \; ; \; Y \mapsto -Y-1.$$

$$P = [X,Y] \quad \longleftrightarrow \quad \bar{P} = [X, -Y-1]$$

$$\begin{array}{ccc} [1,0] & \longleftrightarrow & [1,-1], \\ [2,2] & \longleftrightarrow & [2,-3], \\ [6,14] & \longleftrightarrow & [6,-15], \end{array}$$

etc .

---The "basic symmetry"---



$$X \longmapsto X$$
$$Y \longmapsto -Y-1$$

Using this "symmetry" we can manufacture new solutions of our equation from old ones: given the solution [1,0] we can apply the symmetry to "discover" the solution [1,-1], etc. These discoveries are not too exciting, of course, since the symmetry [X,Y] <----->[X,-Y-1] is so elementary. But are there other geometric properties of the graph of our equation that we can use, to force "old solutions" to somehow lead us to new ones?

Collinear points. I now want to use a geometric property of our curve C which is much subtler than the symmetry we have just discussed.

Any line L *in the* (X,Y)*-plane intersects the curve* C *in at most three points.*

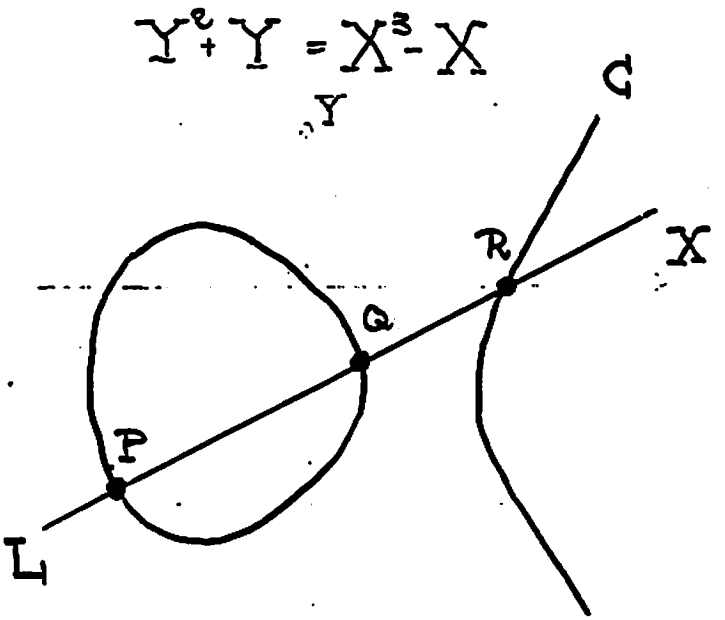
The proof of this is easy: plug into our equation

$$(*) \qquad Y^2+Y = X^3 - X$$

the equation (Y = mX+b) of the line L and solve for X giving you a cubic polynomial in the variable X which can have at most three

solutions !

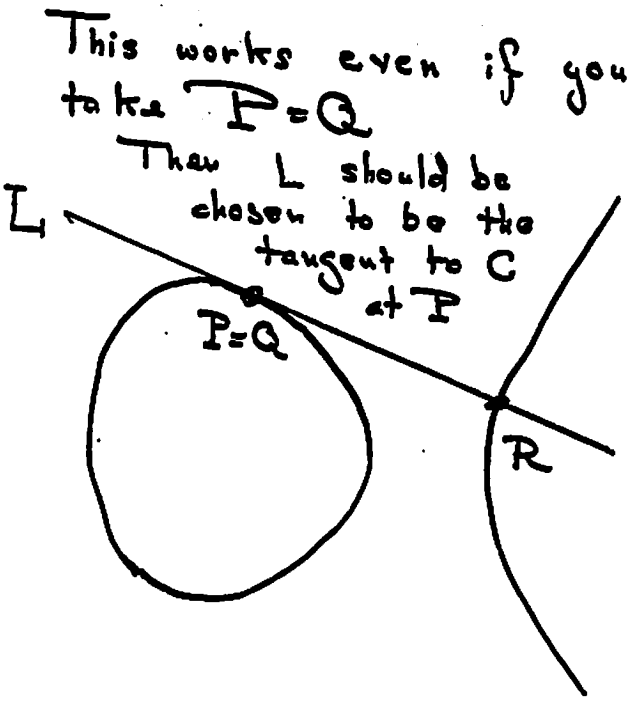
---The "chord" through P and Q gives a "new" point R---



This gives us a strategy which might potentially produce new solutions from old. It has some pitfalls, but I will first formulate it and criticize it only afterwards.

Let $P = [X_1,Y_1]$ and $Q = [X_2,Y_2]$ be two solutions of the equation (*). let L = PQ be the straight line passing through P and Q. For obvious reasons we shall call L the **chord to the curve C passing through P and Q.** Consider the intersection of L and C. There is at most one other intersection point R of L and C. **Solve for R.** This gives a "new" solution $R = [X_3,Y_3]$ of the equation (*). This strategy seems to depend upon having *two distinct solutions* P and Q so as to be able to produce a chord L passing through them. Can we extend this strategy to the case when P=Q; i.e., can we do something silimar starting with only *one* solution? A bit of thought will suggest that YES there is a natural extension of our strategy which allows us to work with one point "P=Q" : take the line L to be simply the tangent line to the curve C at the point P.

---The tangent line through P gives a "new" point R---



This strategy of getting new solutions from old, by the way, is sometimes referred to as the chord-and-tangent process.

Discussion: If we have the points P and Q in hand, we can very easily calculate R. Let us try an example: Take as our two points P = [1,-1] and Q = [2,2]. The line L passing through P and Q has the equation $Y+1 = 3(X-1)$. So, plugging this equation for Y into (*) we get the cubic equation in X:

$$Y^2 + Y = X^3 - X$$

$$(3X-4)^2 + 3X-4 = X^3 - X$$

or:

$$X^3 - 9X^2 + 20X - 12 = 0.$$

Now, X = 1 and X = 2 are solutions of this equation corresponding to the intersection points P and Q on the line L. You can easily then

solve this cubic equation to get its *third* solution, which is

$$X = 6.$$

Then, since $Y = 3X-4$, we get $Y = 14$. That is, our third intersection point R of the curve C with the line L is

$$R = [6,14].$$

The moral, here, is that if we had NOT discovered this solution [6,14] "on our own", we might perfectly well have been led to its existence by this strategy of finding new solutions from old if we had previous gotten the solutions [1,-1] and [2,2]. To put it another way, this is a strategy which forces "old" solutions to "work for us" to possibly produce other solutions.

But there are a few tricky things about this strategy. The first tricky thing, which is minor, is that *for some choices of* P *and* Q *there is no third intersection point* R. This happens if and only if the line L is vertical; this is the same as saying that the X-coordinate of the points P and Q are equal; and this is the same as saying that P and Q are brought to each other by the symmetry of the curve C. The second tricky thing, though, opens up a whole new issue: *Sometimes the point* R *does not have integer coordinates, but only has rational coordinates* . You don't have to go far to run into this. For example, take P = [1,0] and Q = [6,14]. Then the line L passing through P and Q has the awkward equation $5Y = 14(X-1)$ which gives us a denominator of 5 when we solve for Y:

$$Y = (14/5)\cdot(X-1)$$

so that when we proceed as before and finally get the third intersection point of L and C we will find this point R to be:

$$R = [21/25, -56/125].$$

In a word, our strategy does not preserve *integrality* of the solutions that are found, but does preserve *rationality*. So, if we

are going to make any systematic use of this strategy we might be led naturally to consider all *rational solutions* of the equation (*) and to think of the integral solutions (which we were originally after!) as a particular subcollection of rational solutions. You might think that this is a step backwards-- in that there are, very likely, many more rational solutions than there are integral solutions, and therefore our chore is that much harder. We shall follow this path to see where it leads. By a rational solution P to the equation (*) we just mean a pair of rational numbers P =[x,y] that "solve" that equation. If we want to think geometrically we can also call P a **rational point** on C.

Here is the surprising answer to the question: *what are all the rational solutions to* (*) ? Answer: There are infinitely many rational solutions. Nevertheless, don't despair! The magic here is that you can get *all* rational solutions if you start with the single solution P = [0,0] and then proceed to produce "new solutions from old" just by systematically applying the basic symmetry $P \mapsto \bar{P}$ and the chord-and-tangent process to all pairs of points you get along the way. "Nothing will come of nothing" according to King Lear, but as for our problem, the modest "double-zero" solution [0,0] to the equation $Y^2+Y = X^3 - X$ *generates all the infinitely many rational solutions* by the chord-and-tangent process. A minor miracle is that the end-result of all this is amazingly "organized". I will state it as a Theorem.

Theorem: There is an infinity of rational solutions to

$$Y^2 + Y = X^3 - X,$$

and there is a way of "listing" these rational solutions by labelling them in one-one correspondence with the set of all nonzero (positive and negative) integers

$$n \longleftrightarrow P_n = [x_n, y_n]$$

such that this one-one correspondence has these properties:

A. $P_n \longleftrightarrow P_{-n}$ (under the basic symmetry)

B. Three rational points P_n, P_m, and P_r (whose indices are distinct nonzero integers n,m, and r) lie on a straight line L in the (X,Y)-plane if and only if $n+m+r = 0$.

Here is the beginning of this listing for positive values of n (to get the listing for the corresponding negative values just apply the "basic symmetry"; that is, replace the Y-coordinate by -Y-1):

P_1=[0, 0]

P_2=[1,0]

P_3=[-1,-1]

P_4=[2,-3]

P_5=[1/4, -5/8]

P_6=[6, 14]

P_7=[-5/9, 8/27]

P_8=[21/25, -69/125]

P_9= [-20/49, -435/343]

P_{10}=[161/16, -2065/64]

P_{11}=[116/529, -3612/12167]

P_{12}=[1357/841, 28888/24389]

P_{13}=[-3741/3481, -43355/205379]

P_{14}=[18526/16641, -2616119/2146689]

P_{15}=[8385/98596, -28076979/30959144]

P_{16}=[480106/4225, 332513754/274625]

P_{17}=[-239785/2337841, 331948240/3574558889]

P_{18}=[12551561/13608721, -8280062505/50202571769]

P_{19}=[-59997896/67387681, -641260644409/553185473329]

P_{20}=[683916417/264517696, -187844546712 97/4302115807744]

You can check my arithmetic[2], because there are quite a number of miraculous constraints on this list of numbers: they are all solutions to $Y^2+Y = X^3 - X$ but also if you take any pair of distinct integers n,m, by part B of the theorem we formulated, we have that P_n, P_m, and $P_{-(n+m)}$ must lie on a straight line in the (X,Y)-plane. For example, since $3 + 5 + (-8) = 0$, the three points

P_3=[-1,-1], P_5=[1/4, -5/8], and P_{-8}=[21/25, -56/125]

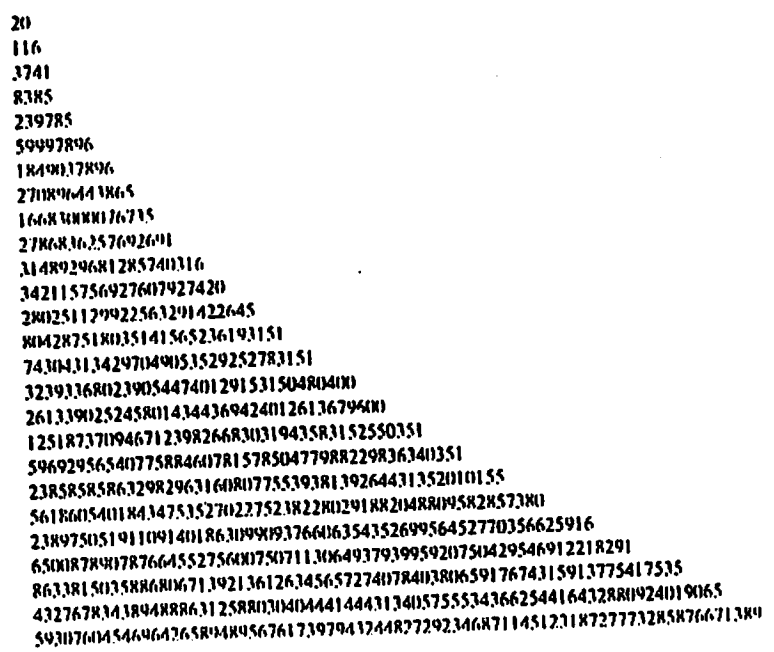
had better be collinear. Otherwise I made a mistake in compiling my list!

I would now like to pause a minute to squint at the list of solutions P_n that is now on the screen. If we are to be as attentive as we possibly can be to the answers that we get to the questions that we ask, there is something about that list-- its general shape-- that should not escape our notice! Do you see a wiggly profile of a *parabola* hidden in it? To bring this out more clearly, let us compactify our data a bit, and consider this slightly more extensive

[2] I confess that this was NOT done by hand: I used the very convenient computer package PARI to compile this

list, where I give only the numerators of the X-coordinates of P_n for even values of n beginning with n = 8 :

---Numerators of the X-coordinates of P_n---

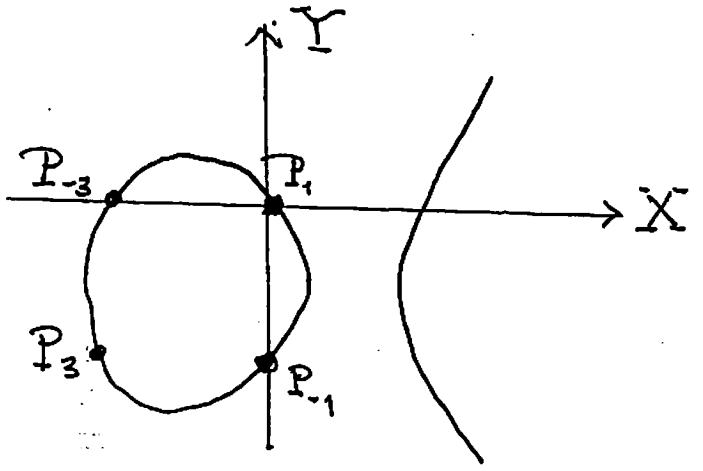


Now, take another look to check out the clean shadow of a parabola formed by the mere digits of our solutions-- this shadow being a vivid indication that the rate of growth of the size of our solutions seems to be following a regular pattern. Can we prove this? Can we "question our answers" so rigorously as to get (heaven help us!) the equations of that parabola? Will doing this lead us to an even deeper understanding of the arithmetic behind our original problem? The answer to all these questions is YES and following their lead would bring you into intimate contact with a good deal of the exciting work in Number Theory that has been taking place in this half century!

APPENDIX: **Rational solutions versus integral solutions.** But we have somehow wandered into the dazzling infinite array of *rational solutions* to our equation $Y^2+Y = X^3 - X$ when, at the outset of our investigation, we had intended to only study the *integral solutions* . We did this because, infinite or not, the rational solutions to the problem have a certain orderly structure that was not in evidence when we focused only on integral solutions. Can we

now go back and pick out the jewels -- the integral solutions-- in our infinite list? The answer here is YES (as was known to Mordell half a century ago) and although my hour talk will not include this, I feel compelled to provide this appendix to give the bare bones of the argument which clinches our problem. There are three facts that you need to calculate. First, we must return to the graph of the of our equation in the (X,Y)-plane and notice that it breaks into two pieces, the oval on the left and the piece "going off the page to infinity" on the right.

---Graph of $Y^2+Y = X^3 - X$ ---



1. Check that the points P_n with *odd index* n lie in the oval on the left while the points P_n with *even index* n lie in the the piece going off to infinity on the right.

2. Check that the only integral points on the oval are the four

integral points.

$P_1=[0, 0]$, $P_{-1}=[0,-1]$, $P_3=[-1,-1]$, and $P_{-3}=[-1,0]$.

3. Check that any prime number dividing the denominators of the X and Y coordinates of P_n also divide the denominator of the X and Y coordinates of P_{2n}.

Supposing that you have done those three chores, and suppose that you know the Theorem we formulated in the lecture and the list of P_n's for small n. Then you are in a position to prove:

Theorem: The only numbers N that are both the product of two consecutive and three consecutive integers are N = 0,6, and 210.

Because: We shall search among the infinite list P_m of rational solutions to $Y^2+Y = X^3 - X$ to see which of these is integral, i.e., which has the property that neither the X nor the Y coordinate has a denominator > 1. If a solution P_m is integral, then its image P_{-m} under "basic symmetry" is also integral, so in searching for all integral solutions we may just try to determine all positive values of m for which P_m is integral. Write $m = 2^e \cdot m_o$ with m_o odd and $e \geq 0$. Use fact **3** to see that for if P_m has integral (X,Y)-coordinates, then for each $j = e-1, e-2, \ldots, 0$, the point $P_{2^j \cdot m_o}$ also has integral (X,Y)-coordinates. In particular, P_{m_o} has integral (X,Y)-coordinates. Since m_o is odd, fact 2 gives us that m_o is either 1 or 3. Now let us consider the case of $m_o=1$ and $m_o=3$ separately.

$m_o=1$: Then m is 2^e and we have seen above that P_{2^j} has integral (X,Y)-coordinates for all $j \leq e$. But, quoting our list,

$$P_{2^3} = P_8 = [21/25, -69/125]$$

does not have integral (X,Y)-coordinates. Therefore $e \leq 2$, and m is either 1,2 or 4.

$m_o=3$: A similar argument, using that

$$P_{22.3} = P_{12}=[1357/841, 28888/24389]$$

is not integral gives us that m is either 3 or 6. □