BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 80, Number 3, May 1974

ON THE INCOMPATIBILITY OF TWO CONJECTURES CONCERNING PRIMES; A DISCUSSION OF THE USE OF COMPUTERS IN ATTACKING A THEORETICAL PROBLEM[1,2]

BY IAN RICHARDS

Introduction. This talk is about the interplay between computers and theoretical research, as experienced by someone who is not a computer expert. The story involves, among other things, a measure of good luck. Several instances of this will emerge in due course, but one example now may give the idea: The speaker and his co-worker, Douglas Hensley, used a computer to seek a certain combinatorial pattern. Our first attempts failed; however the desired patterns did exist, and we eventually found infinitely many of them by theoretical means. If, on the first day, the computer had given us the result we wanted, we probably would have stopped there and missed the further developments.

In line with the purpose of this talk, I intend to be somewhat informal and omit certain details. The main theme is the narrative, which relates how a theoretical argument emerged from a computer search. This begins in Part II, using definitions given in Part I, and the argument itself is sketched at the end of the talk. (A detailed proof of our results will appear in [5].)

As I have mentioned, this work was a collaboration with Douglas Hensley. We were aided in an essential way by William Franta and Richard Franta of our computer sciences department. They programmed a CDC 6400 computer to handle sieving operations on 100,000 points. We used a time-sharing circuit, which proved very helpful since it provided us with instant reinforcement, positive or negative—mostly negative as it turned out.

Our objective was to seek a counterexample to a conjecture. The conjecture involves two functions, the familar prime-counting function $\pi(x)$, and a second function $\rho^*(x)$ related to sieves, which we will define presently. For small values of x, one finds that $\rho^*(x)$ is always smaller than $\pi(x)$, and this inequality was believed to hold generally. Furthermore,

AMS (MOS) subject classifications (1970). Primary 10H15, 10H30, 10-01.

The written version of an invited address given at the Summer Meeting in Missoula, Montana on August 23, 1973. It is not a verbatim rendering of the talk, but is similar in spirit; received by the editors October 1, 1973.

[1] Based on joint work with Douglas Hensley. We were assisted by William Franta and Richard Franta of our computer sciences department.

[2] Partially supported by NSF Grant GP 39050.

the statement that $\rho^*(x) \leqq \pi(x)$, if true, would have important consequences about the distribution of prime numbers. We suspected, however, that the conjecture was false, and we hoped that the computer would help us to locate a counterexample.

In fact the conjecture is false; moreover it is false for all but finitely many x. Although $\rho^*(x) < \pi(x)$ for small x, the functions cross once or several times, and then $\rho^*(x)$ becomes and remains larger than $\pi(x)$. We will prove this by showing that $\rho^*(x) - \pi(x) \to +\infty$.

The main application of our result is to a conjecture of Hardy and Littlewood [**4**] that

$$\pi(x + y) \leqq \pi(x) + \pi(y) \quad \text{for all integers } x, y \geqq 2.$$

We show that this assertion is probably false. More precisely, we prove it incompatible with a second (widely believed) conjecture about "prime k-tuples". The connections between these various number-theoretic statements are spelled out in Part I. That chapter is fairly long; someone mainly interested in the computer side could turn to §1.11 and §1.12 ("The rules of the game") and skip over the rest of it.

It turns out that these questions hinge on the (so far undefined) function $\rho^*(x)$. If it were true that $\rho^*(x) \leqq \pi(x)$, then the Hardy-Littlewood conjecture would follow. But if we can show that $\rho^*(x_0) > \pi(x_0)$ *for even a single value* x_0 *of* x, then we obtain the "incompatibility theorem" mentioned above. This explains the idea behind our computer search.

There was nothing futuristic about our work: our computer was not programmed to "think" or to make logical deductions. It merely calculated in the familiar way. Thus it could only furnish us with particular examples (because it could only handle a finite amount of data). However, as we have seen, our problem was such that even a single counterexample would be interesting. Eventually we found infinitely many of them, by showing that $\rho^*(x) - \pi(x) \to +\infty$, so we were able to dispense with the computer. The machine merely served as a kind of "mechanical scratch pad", helping to put us on the right track.

I am not suggesting that this is the most common way in which computers prove useful in developing theories. There are several aspects of our experience which are slightly paradoxical, and perhaps unusual. For example, since we ultimately proved that $\rho^*(x) - \pi(x) \to +\infty$, you might infer that the computer showed us numerous instances where $\rho^*(x) > \pi(x)$, and that the differences became larger as x increased. Actually no such thing happened. We never found a single case where $\rho^*(x) > \pi(x)$ until we had solved our problem theoretically.

What happened was this: There was a conjecture to the effect that $\rho^*(x) \leqq \pi(x)$. We believed that counterexamples existed (which was

correct), but that the situation was much too complicated to analyze theoretically (which was wrong). So we went to the computer. This gave us something to do, and (more importantly) gave us the courage to keep trying. For there was always the possibility "lurking just around the corner" that the computer would turn up something.

For a long time it didn't. We tried this and that, and sometimes came close, but always failed to produce the desired counterexample. A discussion of what we did, and why it didn't work (using hindsight) is given in Part II. The discussion follows our actual experience: we began with a series of computer experiments, and ended up with an idea—the "midpoint sieve". (The same idea, incidentally, was discovered by Erdös and Selfridge [3], who used it in a different context.)

Only one more idea was needed to solve the problem, but we didn't see it right away, so we went back to the computer. Again the computer data was disappointing, and again it forced us to reconsider our strategy. This is described in Part III. Speaking very loosely, our attempts had been based on the assumption of a certain "randomness." But, much to our chagrin, the desired effect failed to appear. In fact there was a strong regularity principle operating, and things were neither as random *nor as complicated* as we thought. Once we realized this, a theoretical solution followed very quickly.

It turned out that the necessary tools were already available in two well-known theorems. One of these is de la Vallée Poussin's "sharp" form of the prime number theorem [6]; and the other, due to Westzynthius, Erdös, and Rankin [11], [1], [8], asserts the existence of "large" gaps between primes. A sketch of our proof is given in Part IV, using the ideas developed in Parts II and III.

Note. Computing $\rho^*(x)$ exactly is very difficult (cf. §1.7 and §1.8 below). Mostly one tries to get good upper or lower bounds for it. Thus it has been known for some time, from work of Schinzel, Sierpiński, and Selfridge, that $\rho^*(x) \leqq \pi(x)$ for $x \leqq 500$. Recently Warren Stenberg wrote a new computer program which enabled us to show that $\rho^*(x) > \pi(x)$ for $x = 20{,}000$. (Here we are indebted to S. A. Burr for a number of valuable suggestions.) Stenberg's program involves a speeded-up and automated version of the (previously unsuccessful) "midpoint sieve" experiment described in Part III.

Part I. Statement of the problem, and definitions of basic terms.

1.1 Our purpose is to study the number of primes in an interval $(y, y+x]$ of length x, and especially the way that this number behaves when x is held fixed and y approaches infinity. (Here and throughout, x and y denote integers which will generally be $\geqq 2$. By an interval we mean a

sequence of integer points, and its "length" is just the number of these points.) If, as usual, $\pi(x)$ denotes the number of primes $\leqq x$, then the number of primes in $(y, y+x]$ can be written as $\pi(y+x)-\pi(y)$. Hardy and Littlewood [4] conjectured that

$$\text{(A)} \qquad \begin{aligned} &\pi(x+y) \leqq \pi(x)+\pi(y) \quad \text{for } x, y \geqq 2; \quad \text{or equivalently,} \\ &\pi(y+x)-\pi(y) \leqq \pi(x). \end{aligned}$$

In words, (A) asserts that no interval $(y, y+x]$ of length x contains more primes than the first x integers (see Figure 1).

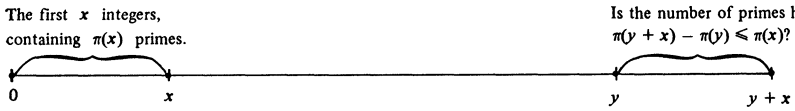


FIGURE 1. The main question.

We will show that the conjecture (A) is probably false. More precisely, we show that (A) is incompatible with the "prime k-tuples conjecture" (B) below. (Curiously, Hardy and Littlewood, in the same paper [4] where they proposed (A), also made the hypothesis (B). We think (B) is true.)

1.2 The statement of (B) is a combinatorial tangle, and it seems better to approach it by examples. You are probably familiar with the "twin primes conjecture" that there are infinitely many prime pairs X, $X+2$. (This has never been proved, of course.) By the way, there is nothing terribly special about the pair X, $X+2$. One can also consider pairs X, $X+4$ or X, $X+50$.

What about X, $X+1$? You probably sense that something is wrong. Two consecutive numbers cannot both be prime, because one of them must be even. (There is a single exception, of course, the pair 2, 3.) Likewise X, $X+2$, $X+4$ is an "impossible" triple for primes, because no matter what value we give to X, one of these numbers must be divisible by three. (Again there is a single exception, 3, 5, 7.)

On the other hand, there seems to be no reason why X, $X+2$, $X+6$ should not all be prime, and we conjecture that this happens infinitely often.

1.3 EXAMPLE. The triple X, $X+2$, $X+6$ takes prime values for $X=11$ and $X=101$, but not for $X=13$ (because in the last case the triple becomes 13, 15, 19, and 15 is not prime).

1.4 We have seen that certain configurations like X, $X+2$ and X, $X+2$, $X+6$ are "possible" for primes, while others like X, $X+1$ and X,

$X+2$, $X+4$ are not. What makes the difference? In each of the "impossible" cases so far examined, there is *a single prime p* which renders the configuration hopeless. (For X, $X+1$ it is two, and for X, $X+2$, $X+4$ it is three.)

More generally, consider a k-tuple $X+b_1, \cdots, X+b_k$, where the b_i are distinct integers. Suppose there is some "small" prime p (like 2 or 3 or 17) such that, no matter what value we give to X, at least one of the values $X+b_1, \cdots, X+b_k$ is divisible by p. Then for "large" values of X, the numbers $X+b_i$ cannot all be prime (because at least one of them must be divisible by p!!!). Thus such a configuration is NOT "possible" for large primes. Now the proper definition practically writes itself (except that instead of "possible" we will say "admissible"). Also, since the constants b_i carry all the information, we may as well consider $b_1, \cdots, b_k$ instead of $X+b_1, \cdots, X+b_k$.

1.5 DEFINITION. A k-tuple of distinct integers $b_1, \cdots, b_k$ (like 0, 2, 6) is *admissible* if, for each prime p, there is some integer X, such that none of the values $X+b_1, \cdots, X+b_k$ is divisible by p. Or equivalently, the condition can be stated:

(*) For each prime p, there is some congruence class (mod p) which contains none of the b_i.

The second alternative (*) is more convenient in practice, and this is the one we shall use.

1.6 EXAMPLE. The sequence 0, 2, 6, 8, 12 is admissible. Why? Well, there are no odd numbers in it, so there is an empty congruence class (mod 2). There are no numbers $\equiv 1$ (mod 3), and no numbers $\equiv 4$ (mod 5). What about 7? This is easy: there are seven congruence classes (mod 7), and only five numbers in our 5-tuple, so there must be an empty class (Dirichlet box principle). Similarly for any prime >5.

REMARK. In general the Dirichlet box principle insures that, to test whether a given k-tuple is admissible, it is only necessary to examine the primes $p \leqq k$.

Now, finally, we can state the prime k-tuples conjecture:

(B) *Let $b_1, \cdots, b_k$ be an admissible k-tuple of integers. Then there exist infinitely many positive integers X for which all of the values $X+b_1, \cdots, X+b_k$ are prime.*

COMPLEMENT TO (B). As we have seen, certain k-tuples of primes like 2, 3 or 3, 5, 7 are not admissible. However, a little thought shows that this can happen only when one of the primes involved is $\leqq k$ (for distinct primes never divide evenly into one another, so the congruence class 0 (mod p) is not occupied by any other prime). In particular, the primes in $(y, y+x]$ form an admissible set provided $y \geqq x$. Thus "admissibility" is obviously a *necessary* condition for the existence of infinitely many

prime k-tuples $X+b_1, \cdots, X+b_k$. The hypothesis (B) asserts that this condition is also sufficient. In other words, "if it could, then it does."

REMARK. A more general conjecture of the same type, in which the functions $X+b_i$ are replaced by arbitrary irreducible polynomials, is known as Schinzel's "Hypothesis H" (cf. [**10**]).

For various "philosophical" reasons, we believe that the k-tuples hypothesis (B) is true. (It has so much freedom—only the existence of infinitely many X is asserted, not even an asymptotic law.)

1.7 DEFINITION. $\rho^*(x)$ denotes the maximum k for which there exists an admissible k-tuple $b_1, \cdots, b_k$ of distinct integers contained in an interval $y<b_i\leqq y+x$ of length x.

Note. It is easy to see that "admissibility" is translation invariant. Therefore, in defining $\rho^*(x)$, we could restrict our attention to any interval of length x, say $(0, x]$. Thus $\rho^*(x)$ is computable in a finite number of steps.

1.8 EXAMPLE. $\rho^*(13)=5$. (Note that by contrast $\pi(13)=6$.) In Example 1.6 we constructed an admissible 5-tuple 0, 2, 6, 8, 12 on the interval [0, 12] of length 13. We leave to the reader the job of showing that there are no admissible 6-tuples.

You may have noticed that I left the hard part to you. This is prophetic of something that happens with larger values of x. We will not actually compute $\rho^*(x)$, because we cannot. (It is as hopeless as computing the optimal strategy in a game of chess.) But any admissible k-tuple that we find contained within $(0, x]$ automatically furnishes a *lower bound* for $\rho^*(x)$. Since our objective will be to prove that $\rho^*(x)>\pi(x)$ for some x, a sufficiently good lower bound is all that we need.

The last definition has now been given. If they seemed interminable, please bear with me. (The trouble is that combinatorics does not yet have such a well-established language as, say, analysis. The underlying geometrical patterns—sequences of dots, say—are simple, but hard to convey in words—see e.g. Figure 2 below.)

The significance of $\rho^*(x)$ is contained in the following:

1.9 PROPOSITION. *Suppose that the k-tuples hypothesis* (B) *holds. Then*: $\rho^*(x)$=*the largest number such that there are infinitely many intervals* $(y, y+x]$ *of length* x *which contain* $\rho^*(x)$ *primes* (*where, as above,* x *and* y *denote integers* $\geqq 2$).

COROLLARY. *Suppose* (B) *holds, and suppose that there is some particular value* x_0 *for which* $\rho^*(x_0)>\pi(x_0)$. *Then* (A) *is false for* $x=x_0$. *Moreover there exist infinitely many* y *for which*

$$\pi(y + x_0) - \pi(y) > \pi(x_0).$$

PROOF. Just a matter of retracing our steps. By definition, $\rho^*(x)$ is the maximum number k of terms belonging to any admissible k-tuple which is contained within an interval of length x. The idea behind "admissibility" is that a configuration $X+b_1, \cdots, X+b_k$ is admissible if it is "possible" or "not trivially disprovable" that this pattern *could* generate primes infinitely often. The hypothesis (B) asserts that "if it could, then it does." This proves the deeper half of (1.9), namely that $\rho^*(x) \leqq$ "the largest number $\cdots$." The reverse inequality, which does not depend on (B), comes from the fact that admissibility is a necessary condition for the k-tuples assertion (complement to (B)). The corollary follows upon observing that $\pi(y+x_0)-\pi(y)$ is just the number of primes in the interval $(y, y+x_0]$. Q.E.D.

Eventually we will show (without any unproved hypotheses) that $\rho^*(x)-\pi(x) \to +\infty$ (cf. Theorem 4.1). We immediately obtain the main result of this paper:

1.10 COROLLARY. *Suppose the k-tuples conjecture* (B) *is true. Then for all sufficiently large x, there exist infinitely many y, such that*

$$\pi(y + x) - \pi(y) > \pi(x).$$

REMARK. Concerning the sizes of x and y: Schinzel and Sierpiński [**10**], [**9**] showed that (A) holds (i.e. (1.10) fails) for $x \leqq 146$, and Selfridge pushed this up to $x \leqq 500$. Warren Stenberg has recently written a program in BASIC, with which we were able to show that $\rho^*(x) > \pi(x)$ for $x = 20{,}000$. Thus "sufficiently large" for x means somewhere between 500 and 20,000. However, the corresponding values of y probably lie far beyond computer range; say around k^k where $k = \pi(20{,}000) = 2262$.

The same examples revisited. In preparation for the computer experiments discussed in Part II, we will take another look at Examples 1.6 and 1.8. Previously we picked a 5-tuple 0, 2, 6, 8, 12 and verified that it was admissible. Let us see how we might discover such a 5-tuple on an interval of length 13. (Later on, the number 13 will be replaced by 100,000.) This requires some preparation; the discussion culminates in Figure 2 below.

1.11 IMPORTANT REMARK. Our computer experiments, and the thought experiments that go with them, involve sieves. We will sieve out congruence classes modulo the ordinary primes $p = 2$, 3, 5, etc., but the points remaining may not be prime, because we may make "crazy" choices, like sieving out all the numbers congruent to 4 (mod 5). The viewpoint we need is a dynamic rather than a static one. We will start with a finite sequence of points (integers), and then erase some of them, and then erase some more, and so on. Once we "erase" or "sieve out" a point, then we regard it as GONE. Thus if we say "the congruence class 4

(mod 5) is empty," we do not mean the absurd statement, "there are no numbers congruent to 4 (mod 5)." Rather we mean that, *among the numbers remaining*, none is congruent to 4 (mod 5). The set of numbers remaining at any stage will be called "the residual set".

1.12 *The rules of the game.* We are going to play a game, and it is necessary to fix in our minds both the "objective" and the "rules" of this game. (As usual, these are in opposition—that is what makes things

	0	1	2	3	4	5	6	7	8	9	10	11	12
0)	•	•	•	•	•	•	•	•	•	•	•	•	•

Initial configuration of 13 points.

	0	1	2	3	4	5	6	7	8	9	10	11	12
1)	•		•		•		•		•		•		•

After first sieving; we erased 1 (mod 2).

	0	1	2	3	4	5	6	7	8	9	10	11	12
2)	•		•				•		•				•

After second sieving; we erased 1 (mod 3).

	0	1	2	3	4	5	6	7	8	9	10	11	12
3)	•		•				•		•				•

After third sieving; we erased 4 (mod 5)
(but this class was already empty).
The residual set = {0, 2, 6, 8, 12}.

0	1	2	3	4	5	6	7	8	9	10	11	12
•		•				•		•				•
	2		2		2		2		2		2	
	3			3			3			3		
				5					5			

The final pattern repeated; this shows which points where sieved out (mod 2), (mod 3), (mod 5). Redundancies are included to show the periodicity.

FIGURE 2a. Sieving operations which produce an admissible 5-tuple contained in the interval [0, 12] of length 13. This shows that $\rho^*(13) \geqslant 5$. Actually $\rho^*(13) = 5$, but proving this requires the examination of other cases, to verify that our procedure is best-possible.

CONGRUENCE CLASS	NUMBER OF POINTS
The count (mod 2), made on line 0 in Figure 2a.	
0	7
1	6 Erase this class; 7 points left.
This produces line 1 in Figure 2a.	
The count (mod 3), made on line 1.	
0	3
1	2 Erase this class; 5 points left.
2	2
This produces line 2.	
The count (mod 5), made on line 2.	
0	1
1	1
2	2
3	1
4	0 Erase this class; 5 points left.
This produces line 3.	

FIGURE 2b. COUNT and ERASE operations which produce the sieving patterns shown in Figure 2a. Our computer print-outs looked almost exactly like this, whereas the information in Figure 2a was hidden in the computer's memory.

interesting.) We start with a finite sequence of consecutive points, and then remove some of them. Our objective is to remove as few as possible. The rules require, however, that for each prime p, one congruence class (mod p) be completely eliminated. The game goes as follows:

(1) We start with an interval of length x (i.e. x consecutive points).

(2) For each prime p, we choose a particular congruence class (mod p). (The choice can be made in any manner whatsoever, but some choices are more intelligent than others.)

(3) Having made the choice in (2), we "sieve out" or "erase" all of the points lying in that class. From now on, these points are gone.

(4) By the Dirichlet box principle, we can stop as soon as the next prime p exceeds the number of points that remain (for then an empty congruence class must exist).

We win if the number of points left exceeds $\pi(x)$. Indeed, our objective

is to find a case where $\rho^*(x) > \pi(x)$; but $\rho^*(x)$ denotes the maximum number of points which can remain after carrying out a sieving process like that described in (1)–(4).

What will be our strategy? Well, the only choice we have comes in Step 2. The simplest strategy is to COUNT how many of the remaining points lie in each congruence class, and then ERASE the class which contains the least. (This is not really very sophisticated; it is like a chess player who thinks only one move ahead. But it is practical.)

Note. The program which the Franta brothers wrote for us had two instructions, ERASE and COUNT. We have just described them above.

Now let us turn to Figure 2. The sieving operations are indicated by the sequence of dots, and the counts are listed below them.

Part II. The computer search.

2.1 Now we begin to play the game described at the end of the preceding chapter. We simply mimic the steps shown in Figure 2, except that we start with a longer sequence of points. Recall that x denotes the length of this sequence. Our objective is to save as many points as possible, while staying within the rules. We win if the number of points left at the end exceeds $\pi(x)$.

First attempt. Here the computer was a human being using pencil and paper. We took $x=1422$, a place where there is a sizable gap in the primes (may as well have $\pi(x)$ small). To win, we needed to finish with more than $\pi(1422)=223$ points left. Our procedure was the same as that shown in Figure 2 (even including the dots), except that now we started with $1422/2=711$ points instead of 13 (the sieving (mod 2) could be done mentally, cutting the number of points in half). At a typical stage, say for the prime 37, we would:

COUNT off $1, 2, 3, \cdots, 37$—$1, 2, 3, \cdots, 37$—etc. and tabulate how many of the remaining points there were in each of the thirty seven congruence classes (mod 37). As you might guess, this was rather tiresome. Also, if we made a mistake and lost our place, the counting had to begin afresh.

Then we would ERASE or cross out all of the points remaining in some particular congruence class (mod 37), choosing the class which contained the least. In case of ties, we took the one nearest the top of the list.

We lost by about 20 points (our goal being to exceed 223).

Second attempt (with the computer, and this time with $x=100{,}000$). We approached the experiment with some anticipation. The program which the Franta brothers had written (it worked perfectly) was put into the machine on a certain Friday, in a part of the memory reserved for frivolities. Then on Monday we got together at one of the teletype

consoles to try it out. However the computer staff routinely erased that part of the memory every Sunday, and our program was gone!

It was put back under a different heading on Tuesday, and on Wednesday we were ready to go. By 2 A.M. Wednesday night, when they shut off the computer, we were quite disappointed. The same thing happened with 100,000 as with 1422: close, but not enough.

Third attempt. We hit on an idea which seemed so good that it had to work. The weak point in our previous attempts was that we were behaving like the chess player who thinks only one move ahead. However, unlike the chess player, we had the option of going back and changing some of our moves. (Also note that our ERASE operations commute; it does not matter whether we first erase all of the numbers $\equiv 1 \pmod 7$ and then those $\equiv 14 \pmod{43}$, or vice-versa. Sieving is, in this respect, less complicated than chess.) The beauty of our plan was that we might improve, but we couldn't do worse, because we could always go back to the move we made before. However there seems to be no virtue in describing the process in more detail, since it didn't work. We improved our score somewhat, but not enough, and pretty soon our position was locked in tight, so that any small changes would only upset it. In other words, we had reached a local maximum.

2.2 *Some speculations.* At this point it might be worthwhile to stop and consider why we were having such bad luck. Perhaps a better question is why we thought we could win in the first place. Recall that our objective was to surpass $\pi(x)$, using a certain type of sieve. But $\pi(x)$, the number of primes $\leqq x$, is also generated by a sieve, the famous sieve of Eratosthenes. To discuss the difference critically, we need one more definition.

2.3 DEFINITION. A sieve or sieving process which eliminates all of the points in one congruence class (mod p) for each prime p is called *hard.* However, if ALL BUT ONE point is removed from the congruence class in question, then the sieve is said to be *soft.*

The sieves we were using above were "hard" sieves—this was forced on us by the rules of our game. By contrast, the familiar sieve of Eratosthenes is a "soft" sieve: it eliminates all multiples of 2 except two itself, all multiples of 3 except three itself, etc. That is why some groups of primes (e.g. 2,3) are not admissible, and also why, for small values of x, $\pi(x) > \rho^*(x)$.

The sieve of Eratosthenes gains an advantage by its softness (the extra points saved). But we also had an advantage in that we could make a lot of choices. We hoped that, when x became large, this freedom would prove decisive.

To see what went wrong, we need two remarks. First, recall that by the prime number theorem, $\pi(x) \sim x/\log x$. Note that this is much larger than

$x^{1/2}$. Secondly, remember that in our game we take the primes in order $p=2, 3, 5, \cdots$, and apply some variant of the "hard" sieve, erasing one congruence class (mod p) for each p. We are permitted to stop only after p becomes larger than the number of points which remain (Dirichlet box principle). Presumably the number of points left at the end will be $\sim\pi(x)$ (slightly larger, we hope). So we stop when $p\sim\pi(x)$. Again I emphasize that $\pi(x)$ is much larger than $x^{1/2}$.

Now unfortunately the sieve of Eratosthenes is very efficient. It has certain "magical" properties which are a consequence of unique factorization. Namely, recall the familiar rule:

To obtain the primes $\leqq x$, sieve out all nontrivial multiples of the primes $p\leqq x^{1/2}$.

Thus the process stops at $x^{1/2}$. By contrast our computer sieves, doing something different with each prime p (say sieving out 1 (mod 2), 0 (mod 3), 4 (mod 5), or whatever) had completely destroyed any pattern like unique factorization. As a result, after we had taken care of the primes $\leqq x^{1/2}$, we would be way ahead, having many more than $\pi(x)$ points left. But then our "scrambled up" sieve would go on killing about $(1/p)$th of the remaining points with each prime p, and since there are a lot of primes between $x^{1/2}$ and $\pi(x)\sim x/\log x$, it would become clear, long before $\pi(x)$ was reached, that we were going to lose. This was frustrating.

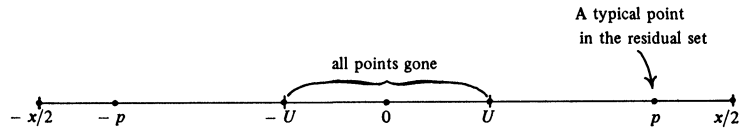
A technical remark. Specialists will recognize, from a result known as Mertens' theorem (cf. [6]), that the sieve of Eratosthenes is more efficient than a "random sieve" by a factor of $e^{\gamma}\cong 1.781$. (The hypothetical "random sieve" is defined as one which takes out $(1/p)$th of the remaining points with each prime p.) Mertens' theorem states:

$$\prod_{p<x}\left[1-\frac{1}{p}\right]\sim\frac{e^{-\gamma}}{\log x}.$$

This "e^{γ} factor" explains why the prime number theorem cannot be proved by purely sieve-theoretic methods.

2.4 *The midpoint sieve.* There is an old saying, "if you can't beat them, join them." Is there a way to turn the sieve of Eratosthenes against itself? (We cannot just take the sieve as it is, because by the rules of our game, we are forced to use a "hard" sieve—see Definition 2.3 above.) Now something that had occurred to us very early (before any computer searches), was the fact that the primes between x and $2x$ form an admissible set, whose size is asymptotic to $\pi(x)$. (Recall that the primes in $(y, y+x]$ form an admissible set provided $y\geqq x$; cf. the "Complement to (B)" preceding Definition 1.7.) However, unfortunately, the density of primes gradually decreases, and $\pi(2x)-\pi(x)$, while asymptotic to $\pi(x)$, is also smaller than it. This should have suggested the obvious step:

Since the primes are thickest near zero, move the origin to the center of the interval! In other words, apply the sieve of Eratosthenes to the symmetric interval $[-x/2, x/2]$. However, by our rules, we have got to use a "hard" sieve: This means that we will take the primes in order, $p=2, 3, 5$, etc., and for each p, remove ALL the multiples of p (including the points $\pm p$). Where do we stop? That turns out to be the crucial question; and so as not to prejudice the issue, I will call the stopping point U. Thus it is understood that we will sieve on all primes $p \leqq U$, and then quit; but the value of U is yet to be determined. In comparing our end result to $\pi(x)$, we will find that moving the origin to the center produces a gain, whereas the use of the "hard" sieve results in a certain loss; we must balance one against the other. For a picture of this, see Figure 3.



Number of points left $= 2\pi(x/2) - 2\pi(U)$.

We want to surpass $\pi(x)$; we reckon our

$$\text{GAIN} = 2\pi(x/2) - \pi(x);$$

(the "gain" represents the advantage obtained by moving the origin to the midpoint of the interval). Our

$$\text{LOSS} = 2\pi(U);$$

(the "loss" is the number of points in the middle which are obliterated by the "hard" sieve–cf. Definition 2.3). Hence

$$\text{GAIN} - \text{LOSS} = (\text{number of points left}) - \pi(x).$$

FIGURE 3. The "midpoint sieve". The sieving will be over all primes $p \leqslant U$, where the stopping-point U is yet to be determined. (Eventually, U will be $x/N \log x$, where N is a constant.) Beginning with the interval $[-x/2, x/2]$, we remove all multiples (positive and negative) of all primes $p \leqslant U$ (the "hard" sieve of Eratosthenes, where the prime itself is not saved). What remains are the primes p between U and $x/2$, each occurring twice as $\pm p$. The primes between $-U$ and $+U$ are gone. (We bypass a couple of technicalities involving one or two points–e.g. the points ± 1, or the endpoints $\pm x/2$.)

2.5 *Gains and losses.* We must compare the gains and losses shown in Figure 3. The gain is $2\pi(x/2)-\pi(x)$. To estimate this, we use de la Vallée Poussin's "sharp" form of the prime number theorem (cf. [**6**]). It follows readily that our

$$(*) \qquad \text{GAIN} \sim (\log 2)[x/(\log x)^2].$$

Sketch of proof. By definition, the "gain"$=2\pi(x/2)-\pi(x)$. Now if you write $\pi(x)\sim x/\log x$, substitute $x/2$ in place of x, and apply a little calculus, you come up with the formula (*). However, strictly speaking, this argument is not correct.

The proper asymptotic law (which must have an error term *smaller* than $x/(\log x)^2$) is:

$$\pi(x) = [x/\log x] + [x/(\log x)^2] + \text{error},$$

where the ratio $[\text{error}]/[x/(\log x)^2]\to 0$. This gives the same result as before, except that now the error is truly negligible.

The loss $2\pi(U)$ depends on the size of U, not yet determined. Obviously we would like to make U as small as possible. However we must end up with an admissible set. The easiest way to achieve admissibility, via the Dirichlet box principle, is to make $U>$ the number of points which remain. Presumably this is about $\pi(x)$ (hopefully a little larger!) so we try $U\sim\pi(x)\sim x/\log x$. Then our loss $\sim 2\pi(x/\log x)$, and by reasoning similar to that in (*) we obtain:

$$(**) \qquad \text{LOSS} \sim 2[x/(\log x)^2].$$

Recalling that our gain was $(\log 2)[x/(\log x)^2]$, and observing that $\log 2<2$, we see that we need something more!

One way to get something more would be to reduce the size of U. For example, suppose we reduce U by some CONSTANT factor $N>2/\log 2$ (e.g. $N=3$). Then our loss would be reduced by the same factor, and we would win. For: our gain would be $(\log 2)[x/(\log x)^2]$; our loss would be $(2/N)[x/(\log x)^2]$.

But now we have to worry about the primes $p>U\sim x/N\log x$. By the rules of our game, before we can quit, there must be an empty congruence class (mod p) for each of these p. Previously this was guaranteed by the box principle; we had fewer points than congruence classes. But now, having lowered the cut-off position U by a factor of $(1/N)$, *we have an average of N points per congruence class* (N is fixed).

Now think of a computer running over all of the p different congruence classes (mod p), and counting how many points lie in each one. The average number of "hits" is N, a constant, whereas the number of "trials" is $p>x/N\log x$. "By the laws of probability" we might expect that, as the

number of trials increases (i.e. as $x \to \infty$), then at least one empty class should appear.

We didn't see any way to prove this, however (although we should have), and so we went back to the computer.

Part III. Computer experiments using the midpoint sieve.

3.1 Now that we had an idea to start from (cf. §2.4 and §2.5) things looked more promising.

Here an aside: At this time we were quite innocent of any special knowledge about computers. The following anecdote will illustrate that remark. We wanted to test our "midpoint sieve". But our program was set to begin at zero and go to 100,000, so in order to move the origin to the midpoint 50,000, it was necessary to do some division. Say for the prime 37, we would divide 37 into 50,000, getting a remainder of 13. Then for the "midpoint sieve" we would erase the congruence class 13 (mod 37), thus hitting the middle point 50,000. Now, thanks to the courtesy of the Franta brothers, we had virtually unlimited access to a large computer. The division algorithm is evidently the sort of thing that a computer can do (in fact, very easily). But it was not in our program, and we had no idea how to do it, so we went out to an all night restaurant, ordered some hamburgers, and did the calculations by hand.

Now to return to our problem. We did an experiment which produced a highly unexpected result. By the way, the question we dealt with is interesting in its own right, and might deserve further study. It is related to the distribution of primes in arithmetic progressions.

First we generated the set of all (positive and negative) primes between $\pm 50{,}000$. (Similar results would have ensued if we had taken the ordinary primes between 0 and 100,000, but we were interested in our "midpoint sieve".) Then we took a particular large prime p; we chose 1511, the first prime above 1500.

I don't want to bore you with a lot of numbers, but we had $2\pi(50{,}000) = 10{,}266$ points left, and we chose the prime $p = 1511$, so we had an average of about $10{,}266/1511 \cong 6.8$ points per congruence class (mod p).

Our objective was to test how "randomly" this set of 10,266 primes is distributed among the 1511 congruence classes (mod 1511). Of course, what we wanted was to find an empty congruence class (that is what "admissibility" is all about). As we suggested at the end of §2.5, if things were "random", then a few empty classes should occur.

It turned out that the distribution is far from random! Specifically, it is too uniform; the values stay much too close to the average. (We are thinking of a Poisson process with mean 6.8 and 1511 trials.) To summarize

the data, as it concerns us: We exclude the class 0 (mod 1511), which is clearly special—it contains just the two primes ± 1511. Otherwise we found no classes with <3 points. In particular, there were no empty congruence classes. (Sorrow!)

3.2 The explanation of this phenomenon is simple: sieving is not a random process! To see why, consider any particular congruence class (mod 1511); more precisely, we want the intersection of that class with the interval $[-50{,}000, 50{,}000]$. What does it look like? Well, we have about 66 ($\cong 100{,}000/1511$) evenly spaced points, exactly 1511 units apart, on an interval of length 100,000 (see Figure 4). [*A key remark*: note that 66 is much smaller than 1511; i.e. the number of points in the congruence class is much smaller than the distance between them.]

Now as far as sieves are concerned, 66 evenly spaced points, with a constant *prime* gap between them, look almost exactly like 66 consecutive integers. For in sieving with the small primes we observe that:

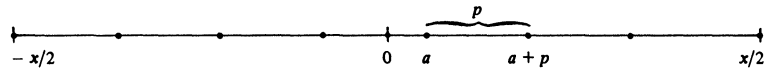


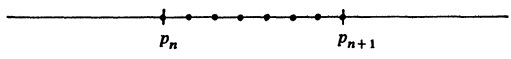
FIGURE 4. The black dots indicate a typical congruence class (mod p), where p is some prime $> x/N \log x$ (see the end of §2.5). We are taking the interval $[-x/2, x/2]$ as our "universe" and ignoring any points lying outside it. The relevant magnitudes are:

$$\text{length of interval} = x,$$
$$\text{distance between points} = p > x/N \log x,$$
$$\text{number of points} \cong x/p < N \log x.$$

Note that the number of points, $N \log x$, is much smaller than the distance between them, $x/N \log x$. (N is constant.)

[In the numerical example at the end of Part III, $x = 100\ 000$, $p = 1511$, and the number of points per congruence class $\cong 100\ 000/1511 \cong 66$.]

Nothing in this figure is very deep—mainly we want to illustrate the analogy between the sequence of $N \log x$ widely separated points shown above, and a sequence of $N \log x$ consecutive integers. The latter appears as a "gap between primes" in the Westzynthius, Erdös, Rankin Theorem, (*) in §4.4.



The prime 2 *eliminates every other point, the prime* 3 *eliminates every third point, and so on, regardless of whether we are dealing with consecutive integers or points* 1511 *units apart.*

Hence the nonrandomness! Now we can dispense with the computer; the material is at hand for a theoretical solution.

Part IV. Proof of the main result. Recall that by Definition 1.7, $\rho^*(x)$ is the maximum number k of terms belonging to any admissible k-tuple which is contained within an interval of length x. "Admissibility" is defined in §1.5.

4.1 THEOREM. $\rho^*(x)-\pi(x)\to+\infty$ *as* $x\to\infty$. *The difference is greater than* $(\log 2-\varepsilon)[x/(\log x)^2]$, *where* ε *denotes a function* $\varepsilon(x)$ *which goes to zero as* $x\to\infty$.

PROOF. (Sketch; for more details see [5].) We will construct a set, contained in the interval $[-x/2, x/2]$, which has nearly $2\pi(x/2)$ points, and then we will show that this set is "admissible".

4.2 *The basic construction.* Fix any constant $N>2/\log 2$. Take the interval $[-x/2, x/2]$, where x is "large enough to make the following arguments work." Eliminate all multiples (positive and negative) of all primes $p\leqq x/N\log x$. What remains of the original interval will be called the *residual set*. Now we will show that, for large x:

(i) The residual set has more than $\pi(x)$ points, by an amount asymptotic to: $[\log 2-(2/N)][x/(\log x)^2]$.

(ii) The residual set is admissible in the sense of Definition 1.5.

PROOF OF (i). This has already been proved in the section entitled "Gains and losses", §2.5 at the end of Part II. There we introduced the constant N, and showed that this gives (i). However we also noted that it causes a difficulty with (ii).

PROOF OF (ii). Recall that a set is admissible if, for EVERY prime p, there is some congruence class (mod p) which does not intersect that set. For the primes $p\leqq x/N\log x$, this is assured in our case, since we have sieved out the congruence classes 0 (mod p). The problem comes with the primes $p>x/N\log x$ which we have left untouched.

4.3 As so often happens, our problem is two thirds solved once we look at it the right way. (Here see Figure 4.) First we will consider the interval $[-x/2, x/2]$ as our "universe" and ignore any points lying outside it. Now take any prime $p>x/N\log x$, and consider a typical congruence class (mod p). Our interval has length x, and the points in the congruence class are $p>x/N\log x$ units apart. Thus each congruence class (mod p) intersects $[-x/2, x/2]$ in less than $N\log x$ points. (Note that this isn't very many—$\log x$ is small compared to x.) Furthermore, a congruence class—any congruence class, to any modulus—is a sequence of equally

spaced points. This is the viewpoint we want: we are carrying out sieving operations on a "short" sequence of $< N \log x$ equally spaced points.

4.4 Now we carry over some results of Westzynthius, Erdös, and Rankin [**11**], [**1**], [**8**] dating back to the 1930's. These results concern unusually large gaps between primes; they assert (we need only the weakest form):

$$\limsup_{n\to\infty} [p_{n+1} - p_n]/\log p_n = \infty \qquad (p_n = \text{the } n\text{th prime}).$$

Note that by the prime number theorem, the average gap is $\log p_n$; the W.E.R. theorem says that occasionally the gap is larger, by a factor going to infinity. We will not give the proof; our paper is incomplete at this point. For the missing steps, see [**8**] or [**5**]. However we will state the main idea behind the proof, and then use that idea in a slightly different context.

In the first place, a gap between primes is the same thing as a string of composite numbers. Westzynthius, Erdös, and Rankin construct a certain sequence of consecutive integers, and prove these integers composite by showing that they have small prime factors. As you might guess, their argument is based on a sieve.

REMARK (see also Figure 5). The W.E.R. results are based on a "killing

The "gap" between the primes 113 and 127.

114	115	116	117	118	119	120	121	122	123	124	125	126
•	•	•	•	•	•	•	•	•	•	•	•	•
2		2		2		2		2		2		2
3			3			3			3			3
	5					5					5	
					7							7
							11					

ABOVE: Small divisors which eliminate these numbers from the set of primes. Redundancies are included to show the periodicity.

FIGURE 5. The "killing sieve". The 13 composite numbers from 114 to 126 inclusive are killed by the primes $\leqslant 11$. There is no more efficient way to sieve out 13 consecutive numbers.

sieve"—one that removes all points. For their "string of composite numbers" is just a sequence of consecutive integers, all of which are removed by the sieve of Eratosthenes. By contrast, our paper is ultimately concerned with a "saving sieve"—one that preserves as many points as

possible. The two sieves are closely related. For, going back to "the rules of the game" (§1.12), we recall that our objective was to save points, whereas the rules require an empty congruence class (mod p) for each prime p. With those $p \leqq x/N \log x$, we construct this class explicitly, by sieving out the multiples of p (i.e. the class 0 (mod p)). But for $p > x/N \log x$, we intend to show that there is some congruence class already "killed", so no further sieving is necessary.

What Westzynthius, Erdös, and Rankin prove is:

(*) Fix any constant $N>0$. Then on any interval of length x (provided x is large enough), there exist SEQUENCES OF $N \log x$ CONSECUTIVE INTEGERS, such that each integer in the sequence is divisible by some "small" prime $p_0 < \log x$.

4.5 Now go back to the end of §4.3 (see also Figure 4). Note that in both §4.3 and (*) we have a situation involving $N \log x$ equally spaced points. Namely:

In (*) we have $N \log x$ consecutive integers.

In §4.3 we have a congruence class (mod p), which is a sequence of $< N \log x$ equally spaced points, p units apart. [Recall that we are looking at an interval of length x, and that $p > x/N \log x$.]

One more analogy: Both the Westzynthius, Erdös, and Rankin result (*) and our problem involve choices. In (*), there is a distinguished subinterval of $N \log x$ consecutive integers contained within a larger interval of length x. In our case, again starting with an interval of length x, we can choose any one of the $p > x/N \log x$ congruence classes (mod p). Each congruence class consists of $< N \log x$ equally spaced points.

The only difference is that in (*) the points are close together, whereas in our problem they are far apart. But we observed at the end of Part III that "consecutive integers" and "equally spaced points" do not look very different to a sieve. Thus it is not surprising that, by a little playing with the Chinese remainder theorem, we can transform the Westzynthius, Erdös, and Rankin result (*) to read:

(**) if $p > x/N \log x$, and x is sufficiently large, then there exists a CONGRUENCE CLASS (mod p), each of whose elements (in the interval $[-x/2, x/2]$) is divisible by some "small" prime $p_0 < \log x$.

[Of course we have not proved (**), but merely indicated its plausibility by analogy with the known result (*) of Westzynthius, Erdös, and Rankin.]

Now we are finished. For our objective was to find an "empty" congruence class (mod p), i.e. a class containing no points of the residual set. But the class given by (**) certainly fits that description, since it contains only multiples of the small primes $p_0 < \log x$, whereas primes up to $x/N \log x$ have been sieved out.

Further results. The paper [**5**] gives a survey of the present state of knowledge about $\pi(y+x)-\pi(y)$. For instance, there are upper bound

results due to Montgomery and others which complement the lower bounds obtained by us. Schinzel has found an extension of our theorem, requiring a second unproved hypothesis in addition to (B), but still suggesting very strongly that our results are not best-possible [**5**, §4]. One of the main unsolved problems in this area is whether $\rho^*(x) \sim \pi(x)$. The most that we know is that $\rho^*(x) \leqq 2\pi(x)$ for all $x \geqq 2$ (Montgomery and Vaughan), and that $\rho^*(x) > \pi(x)$ for large x (Theorem 4.1 above). A complete proof of Theorem 4.1 appears in [**5**, §2].

In conclusion. Although a few details have been omitted--specifically the proof of (**) at the end of Part IV--I have faithfully recorded the turning points in our argument. These were the midpoint sieve, developed in §2.4 and §2.5, and the periodic nature of congruence classes, remarked at the end of Part III and again in §4.3. The proof of (**) involves technical arguments which look formidable at first glance but are familiar to specialists. It was the *connection* between (**) and our problem which had remained hidden. Concerning the title of this talk: as our friend Warren Stenberg put it--

We could have solved the problem without the computer, but we probably wouldn't have.

References

1. P. Erdös, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.

2. ———, *Some unsolved problems*, Michigan Math. J. **4** (1957), 291–300. MR **20** #5157.

3. P. Erdös and J. L. Selfridge, *Complete prime subsets of consecutive integers*, Proc. Manitoba Conf. on Numerical Mathematics, University of Manitoba, Winnepeg, 1971, pp. 1–14.

4. G. H. Hardy and J. E. Littlewood, *Some problems of "partitio numerorum". III. On the expression of a number as a sum of primes*, Acta Math. **44** (1923), 1–70.

5. D. Hensley and I. Richards, *Primes in intervals*, Acta Arith. **25** (1974), 375–391.

6. A. E. Ingham, *The distribution of prime numbers*, Cambridge Tracts in Math. and Math. Phys., no. 30, Stechert-Hafner, New York, 1964. MR **32** #2391.

7. D. H. Lehmer, *Computer technology applied to the theory of numbers*, Studies in Number Theory, Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.), 1969, pp. 117–151. MR **40** #84.

8. R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.

9. A. Schinzel, *Remarks on the paper 'Sur certaines hypothèses concernant les nombres premiers'*, Acta Arith. **7** (1961/62), 1–8. MR **24** #A70.

10. A. Schinzel and W. Sierpiński, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185–208; erratum **5** (1959), 259. MR **21** #4936.

11. E. Westzynthius, *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Comm. Phys. Math. Helsingfors (5) **25** (1931), 1–37.

Department of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455