A WINNING STRATEGY AT TAXMAN®

DOUGLAS HENSLEY
Texas A & M University, College Station, TX 77843-3368

(Submitted October 1986)

Taxman® is an educational computer game, brought out by the Minnesota Educational Consortium. Starting from an initial set, which in the standard game is

$$S_1 = \{1, 2, \ldots, n\},$$

the player chooses successive integers $k_1, k_2, \ldots$. After each choice k_j, k_j and its divisors in S_j are deleted to form S_{j+1}. The player's score is increased by k_j and the computer's by the sum of all the deleted proper divisors. It is illegal to choose $k \in S_j$ if k has no proper divisor in S_j. Initially, any k except 1 may be chosen in the standard game, since that k has at least the proper divisor $1 \in S_1$. As play continues, the number of legal choices dwindles. Whenever the player has no legal move, the computer scores the sum of the remaining elements and the game is over. The objective is to have a higher score than the computer at the end.

Play can be described by listing the integers chosen in the order they were picked. For instance, with $n = 10$, we might play (10, 9, 8). The monitor would show, successively,

		YOU	ME
{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}	,	0,	0
{3, 4, 6, 7, 8, 9}	,	10,	8
{4, 6, 7, 8}	,	19,	11
{6, 7}	,	27,	15
GAME OVER	,	27,	28.

We lost. We could have won if we had picked 7 first. The computer would have deleted 7 (for us) and 1 (for itself) to give $S_2 = \{2, 3, 4, 5, 6, 8, 9, 10\}$. After that we could still have chosen 10, 9, and 8, or better still, 9, 6, 8, and 10. In general, we should begin play by choosing the largest prime $p \leqslant n$. Aside from our choice, only 1 will be deleted, and it is deleted on any first move. However, for large n there are $\approx (1/2)n^2$ points at stake, and this tactic makes at most an n point difference. Let $f(n)$ denote the best possible score for the player on $\{1, 2, \ldots, n\}$. It is natural to conjecture that

$$\lim_{n\to\infty} f(n)\Big/\left(\frac{1}{2}n^2\right) = C$$

exists. If so, and if $C > 1/2$, then the player can win for all sufficiently large n.

In fact, we have the following theorem.

Theorem: $\lim\limits_{n\to\infty} f(n)\Big/\left(\frac{1}{2}n^2\right) = C$ exists, and $\frac{1}{2} < C < \frac{3}{4}$.

From this it follows that the player can win for n sufficiently large. On the basis of the proof we give, "sufficiently large" may be very large indeed. Yet a little experimentation strongly suggests that in fact the player can win for $n \geqslant 4$. Resolving the question of how large n has to be is simple, in principle. Suppose our theoretical argument shows $f(n) > (1/4)(n^2 + n)$ for $n > N$. We have only to exhibit a winning line of play for all n, $4 \leqslant n \leqslant N$, to show the player wins for any $n \geqslant 4$. Unfortunately, the calculations will be lengthy unless the theoretical argument is greatly sharpened, reducing N to tractable size. (I obtained $N = 6{,}000{,}000$.)

The idea of the asymptotically winning strategy is to divide and conquer, by partitioning the game into subgames playable separately on certain nonstandard initial sets. We select a prime p and let

D_p denote $\{d\colon$ if $q|d$ and q is prime, then $q \leqslant p\}$,

$$A_p = \prod_{\substack{q \leqslant p \\ q \text{ prime}}} q \quad \text{and} \quad B_p = \prod_{\substack{q \leqslant p \\ q \text{ prime}}} (1 - 1/q).$$

Thus, $D_3 = \{1, 2, 3, 4, 6, 8, 9, 12, 16, \ldots\}$, $A_3 = 6$, and $B_3 = 1/3$. Next, for the chosen p, we partition $\{1, 2, \ldots, n\}$ into sets

$$N_{k,p}(n) = \{kd\colon d \in D_p \text{ and } kd \leqslant n\}$$

for k relatively prime to A_p. Thus, with $p = 3$, $\{1, 2, \ldots, 40\}$ partitions as

$\{1, 2, 3, 4, 6, 8, 9, 12, 16, 18, 24, 27, 32, 36\}$, $\{5, 10, 15, 20, 30, 40\}$
$\{7, 14, 21, 28\}$, $\{11, 22, 33\}$, $\{13, 26, 39\}$, $\{17, 34\}$, $\{19, 38\}$

and some singletons. Any time we choose a number in $N_{1,p}(n)$, only elements of $N_{1,p}(n)$ are deleted. In general,

However we play on $N_{k,p}(n)$, *the sets* $N_{k',p}(n)$ for $k' > k$ *are undisturbed.* (1)

Let $f_p(x)$ denote the best score possible for the player if the (nonstandard) initial set is $\{d \in D_p\colon\ 1 \leqslant d \leqslant x\}$. Thus, $f_p(x)$ is defined for real x, but only changes at elements of D_p; e.g., $f_3(5) = 7$, because on $[1, 5] \cap D_3 =$

$\{1, 2, 3, 4\}$, the best play is to take 3 and then 4. Similarly, $f_3(36) = 144$, taking 3, 4, 27, 18, 36, 24, and 32, starting from $[1, 36] \cap D_3$.

The best score possible on $N_{k,p}(n)$ is clearly $kf_p(n/k)$. In view of (1), then, if we play on $N_{k,p}(n)$ in order of increasing k, we get

$$\sum_{k \leqslant n}^{*} kf_p(n/k),$$

where $\sum^*$ denotes summation only over k relatively prime to A_p. This score is a lower bound for $f(n)$, that is,

$$f(n) \geqslant \sum_{k \leqslant n}^{*} kf_p(n/k). \tag{2}$$

In our example $n = 40$, the same line of play is applied to $\{11, 22, 33\}$ and $\{13, 26, 39\}$, and from these we score $11f_3(3)$ and $13f_3(3)$, respectively. In general, grouping partition pieces having the same number of elements puts (2) into the form

$$f(n) \geqslant \sum_{\substack{j \leqslant n \\ j \in D_p}} f_p(j) \sum_{\frac{n}{j'} < k \leqslant \frac{n}{j}}^{*} k, \tag{3}$$

where j' denotes the next element of D_p after j.

Let us now temporarily put aside rigor and look ahead to the answer. If $B_p = b/A_p$, then b is an integer, and of any A_p consecutive integers, b of them are relatively prime to A_p. Thus, the inner sum in (3) is the sum of, roughly, $B_p\left(\frac{n}{j} - \frac{n}{j'}\right)$ integers, with an average value of about $\frac{1}{2}\left(\frac{n}{j} + \frac{n}{j'}\right)$. This suggests something like

$$f(n) \geqslant \frac{1}{2}n^2 B_p \sum_{\substack{j \leqslant n \\ j \in D_p}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right). \tag{3'}$$

Happily, essentially the same sum as in (3) provides an upper bound for $f(n)$.

Suppose we choose p prime and then play a game on $\{1, 2, \ldots, n\}$. For each integer $m \leqslant n$ we pick, we note which is the largest proper divisor of m in play at the time, and call it $t(m)$. Distinct m's have distinct $t(m)$'s, since $t(m)$ is deleted when we pick m. We separate the m we pick into two sets:

$$M_1 = \{m\colon t(m) < n/p\} \quad \text{and} \quad M_2 = \{m\colon t(m) \geqslant n/p\}.$$

Clearly, M_1 has fewer than n/p elements, so our score from M_1 is less than n^2/p. To bound from above our score from M_2, we need a lemma.

Lemma 1: Suppose k is relatively prime to A_p, $d \in D_p$, and $kd \in M_2$. Then

$$k \mid t(kd).$$

Proof: If $k \nmid t(kd)$, then some prime $q > p$ divides k to a higher power than it does $t(kd)$. But then

$$kd/t(kd) \geqslant q > p,$$

in contradiction to the assumption $t(kd) \geqslant n/p$. ■

From Lemma 1, we claim that, for $k \leqslant n$ and relatively prime to A_p,

$$\sum_{\substack{d \in D_p \\ kd \in M_2}} kd \leqslant kf_p(n/k). \tag{4}$$

For consider the sequence of moves in the standard game we just played, but restricted to those moves which chose a number of the form kd, with $d \in D_p$ and $kd \in M_2$. We can map this sequence of moves onto a shadow game played on the initial set $D_p \cap \{1, 2, \ldots, [n/k]\}$. The image of a choice of kd in the real game is the choice of d in the shadow game. This d will be a legal move. First, $k^{-1}t(kd)$ is a proper divisor of d, since $t(kd)$ was a proper divisor of kd, and since, from Lemma 1, $k^{-1}t(kd)$ is an integer. Second, since $t(kd)$ had not yet been deleted at the time kd was chosen in the standard game, no multiple kd' of $t(kd)$ had yet been chosen in that game. Thus, in the shadow game, no multiple d' of $k^{-1}t(kd)$ can yet have been chosen. Therefore, $k^{-1}t(kd)$ must still be in play in the shadow game and available as an as yet undeleted proper divisor of d. By its definition, the sum of the numbers d so chosen in the shadow game is less than or equal to $f_p(n/k)$, and (4) follows on multiplication by k.

Summing (4) over k and using our observation about M_1 now gives

$$\sum_{m \in M_1 \cup M_2} m \leqslant n^2/p + \sum_{k \leqslant n}^{*} kf(n/k), \tag{5}$$

and since this holds even for best play, we can group k's as before and get

$$f(n) \leqslant n^2/p + \sum_{\substack{j \leqslant n \\ j \in D_p}} f_p(j) \sum_{\frac{n}{j'} < k \leqslant \frac{n}{j}}^{*} k. \tag{6}$$

The analog of (3′) is then

$$f(n) \leqslant \frac{1}{2}n^2\left(\frac{2}{p} + B_p \sum_{\substack{j \leqslant n \\ j \in D_p}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)\right). \tag{6′}$$

Now assuming that the sum here is convergent (and it is, as we shall prove later), (3′) and (6′) converge to give

$$\lim_{n \to \infty} \frac{f(n)}{(1/2)n^2} = \lim_{p \to \infty} B_p \sum_{j \in D_p} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) = C. \tag{7}$$

The path now splits. We should like to have some notion of the value of C, and the demands of rigor must be met. First, let us work on C.

In principle we have only to pick p large, calculate $f_p(j)$ for enough terms that the "tail" of

$$\sum_{D_p} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

is less than $1/p$, and we shall get C to within an error on the order of $1/p$. The catch is that it is hard to find $f_p(j)$ for large j and large p.

A crude upper bound is not so hard. Any odd numbers m between $n/2$ and n that are picked have odd $t(m)$ between 1 and $n/3$. There are, thus, $\leqslant (n+3)/6$ such m. We can pick in all no more than $(1/2)n$ numbers. The sum of a set of $\leqslant n/2$ numbers, all $\leqslant n$ and containing at most $(n+3)/6$ odd numbers between $n/2$ and n, is at most $(35/96)n^2 + 0(n)$.

Thus, $C \leqslant 35/48 < 3/4$. The proof that $C > 1/2$ is more difficult.

We choose $p = 5$ and calculate $f_5(j)$ for $1 \leqslant j \leqslant 36$, and then a lower bound for $f_5(j)$ for $40 \leqslant j \leqslant 200$. It turns out that

$$\sum_{\substack{j \in D_5 \\ j \leqslant 200}} f_5(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) > 1.9 \quad \text{and} \quad B_5 = \frac{4}{15}.$$

Now, $$C \geqslant B_5 \sum_{j \in D_5} f_5(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) > \frac{4}{15} \cdot \frac{19}{10} > \frac{1}{2}.$$

[See the table of $f_5(j)$, 1 to 36, and the lower bound, 40 to 200.] More extensive calculations with $p = 7$ suggest that in fact $C > .56$. Before proceeding to the problem of justifying (3′) and (6′) (which are not claimed to hold verbatim), it would be well to spell out the winning strategy.

(A) Partition $\{1, 2, \ldots, n\}$ into sets of the form $N_{k,5}(n) = \{kd\colon d \in D_5, d \leqslant n/k\}$ with k relatively prime to 30.

(B) Discard $N_{k,5}(n)$ if $n/k > 200$. Make no attempt to score from these k.

(C) For all k relatively prime to 30 and satisfying $(n/200) \leqslant k \leqslant n$, play $N_{k,5}(n)$ as instructed by the table. Start with smaller values of k and work up.

This will win if n is large enough. For lesser n, we might do well to go ahead and play the $N_{k,5}(n)$ for small k by ear, starting with $k = 1$. And, of course, first pick the largest prime.

We now justify (3′) and (6′) and show that

$$\sum_{D_p} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

is convergent. The "0" notation will be helpful from this point on. We say

$\phi_1(n) = 0(\phi_2(n))$ if there exists $C_1 > 0$ such that $|\phi_1(n)| \leqslant C_1\phi_2(n)$ for all n. A subscript 0_p denotes that for each p such a constant C_p exists.

Table A. $f_5(n)$ for $n \in D_5$, $n \leqslant 36$

n	$f_5(n)$	Moves
1	0	none
2	2	(2)
3	3	(3)
4	7	(3, 4)
5	9	(5, 4)
6	15	(5, 4, 6)
8	19	(5, 6, 8)
9	28	(5, 9, 6, 8)
10	33	(9, 6, 10, 8)
12	44	(5, 9, 10, 8, 12)
15	54	(9, 15, 10, 8, 12)
16	62	(9, 15, 10, 12, 16)
18	80	(9, 15, 20, 18, 12, 16)
20	96	(5, 15, 10, 20, 12, 18, 16)
24	112	(9, 15, 10, 18, 20, 16, 24)
25	128	(25, 15, 10, 20, 16, 18, 24)
27	155	(25, 15, 27, 10, 18, 20, 16, 24)
30	177	(2, 25, 15, 27, 18, 30, 20, 16, 24)
32	193	(2, 25,15, 27, 18, 30, 20, 24, 32)
36	219	(3, 4, 25, 27, 18, 36, 24, 20, 30, 32)

Table B. The lower bound for $f_5(n)$ given here comes from first playing the odd numbers by hand, then taking $2f_5(n/2)$ for our score on the evens.

n,	$f_5(n) \geqslant$,	Moves	n,	$f_5(n) \geqslant$,	Moves
40,	259,	(25, 15, 27)	100,	941	
45,	292,	(3, 25, 27, 45)	108,	1049	
48,	324		120,	1137	
50,	356		125,	1239,	(5, 125, 75, 45, 81)
54,	410		128,	1303	
60,	454		135,	1402,	(5, 9, 81, 125, 75, 135)
64,	486		144,	1506	
72,	538		150,	1610	
75,	590,	(5, 27, 45, 75)	160,	1770	
80,	670		162,	1924	
81,	747,	(3, 25, 75, 45, 81)	180,	2056	
90,	813		192,	2184	
96,	877		200,	2312	

Lemma 2: For $0 < x < y$,

$$\sum_{x<k\leqslant y}^{*} k = \frac{1}{2}B_p(y^2 - x^2) + 0(A_p^2) + 0(yA_p).$$

Proof: Consider the set R_p of reduced residues mod A_p that are relatively prime to A_p. R_p has A_pB_p elements. For each $r \in R_p$, the arithmetic progression $(r, r + A_p, r + 2A_p, \ldots)$ intersects the interval (x, y) in either $[(y - x)/A_p]$ or $1 + [(y - x)/A_p]$ points, whose average lies between $\frac{1}{2}(x + y - A_p)$ and $\frac{1}{2}(x + y + A_p)$ if there are any. Thus, for $r \in R_p$,

$$\sum_{\substack{k = r + jA_p \\ x < k < y}} k = \left(\frac{x + y}{2} + O(A_p)\right)\left(\frac{y - x}{A_p} + O(1)\right) = \frac{y^2 - x^2}{2A_p} + O(y) + O(A_p). \tag{8}$$

Now, summing over the A_pB_p elements of R_p gives

$$\sum_{x < k < y}^{*} k = \frac{1}{2}B_p(y^2 - x^2) + O(yA_p) + O(A_p^2). \blacksquare \tag{9}$$

Remark: We could get much sharper estimates here from the literature on sieves. The quantity estimated in (9) is a weighted count of how many numbers survive sifting by the small primes $q \leqslant p$. See [2] for a readable introduction to sieves.

From Lemma 2, and from (3),

$$\sum_{k \leqslant n} kf_p(n/k) = \frac{1}{2}n^2B_p \sum_{\substack{j \in D_p \\ j \leqslant n}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) + O_p\left(n \sum_{\substack{j \in D_p \\ j \leqslant n}} \frac{1}{j} f_p(j)\right). \tag{10}$$

Now, let $g_p(j)$ be the total number of points at stake in $D_p \cap [1, j]$, that is,

$$g_p(j) = \sum_{\substack{d \leqslant j \\ d \in D_p}} d. \tag{11}$$

Then $f_p(j) < g_p(j)$, so (10) holds with g_p in place of f_p in the error term. Now,

$$\frac{1}{j}g_p(j) \leqslant \sum_{\substack{d \in D_p \\ d \leqslant j}} 1 = \Psi(j, p).$$

The counting function $\Psi(x, y)$ of integers $\leqslant x$ composed exclusively of primes $\leqslant y$ has been the topic of numerous studies over the past fifty years. For an elementary but surprisingly good estimate, see [1].

Here, because we are not trying to see how small we can take n with a given p, a simple estimate will do.

Lemma 3: $\Psi(x, p) = O(\log x)^p$.

Proof: There are $\leqslant \left[\frac{\log x}{\log 2}\right] + 1$ possible values for the number of powers of 2 in a number $\leqslant x$, $\left[\frac{\log x}{\log 3}\right] + 1$ possibilities for the number of powers of 3, ..., and there are clearly fewer than p primes $\leqslant p$. $\blacksquare$

Thus, from Lemma 3,

$$\sum_{\substack{j \in D_p \\ j \leqslant n}} \frac{1}{j} g_p(j) = O(\log n)^{2p}, \tag{12}$$

since the sum in (12) has $\Psi(n, p)$ terms. Thus,

$$\sum_{k \leqslant n}^{*} k f_p(n/k) = \frac{1}{2} n^2 B_p \sum_{\substack{j \in D_p \\ j \leqslant n}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) + O_p(n(\log n)^{2p}). \tag{13}$$

Our other unfinished business is to show that

$$\sum_{\substack{j \in D_p \\ j \leqslant n}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

is convergent. For purposes of computation, some estimate of the *rate* of convergence would also be helpful—how many terms must we take to bring the partial sum to within ε of its limit?

Convergence of

$$\sum_{\substack{j \in D_p \\ j \leqslant n}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

is simple. Since $f_p(j) < g_p(j)$, we need only prove

$$\sum_{D_p} g_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

convergent. It is, to $1/B_p$. ■

Proof:
$$\sum_{D_p} g_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) = \sum_{j \in D_p}\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)\sum_{\substack{d \in D_p \\ d \leqslant j}} d = \sum_{d \in D_p} d \sum_{\substack{j \in D_p \\ j \geqslant d}}\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$
$$= \sum_{d \in D_p} d/d^2 = \sum_{d \in D_p} 1/d = \prod_{\substack{q \leqslant p \\ q \text{ prime}}} (1 - 1/q)^{-1} = 1/B_p.$$

Now, for any fixed p, if

$$\sum_{\substack{j \in D_p \\ j \leqslant n}} g_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

is within ε of $1/B_p$, then

$$\sum_{\substack{j \in D_p \\ j > n}} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) < \varepsilon. \tag{14}$$

But how does n in (14) depend on ε and p? Here is an estimate—the technique is taken from probabilistic number theory, and we omit the proof.

$$\sum_{\substack{j \in D_p \\ j \geqslant x}} g_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right) = O\left(\frac{1}{x}(\log x)^p\right) + O\left(\frac{(\log p)^2}{\log x}\right).$$

REFERENCES

1. N. G. de Bruijn. "On the Number of Positive Integers $\leqslant x$ and Free of Prime Factors $> y$ II. *Nederl. Akad. Wetensch. Proc.* Ser. A 69 = *Indag. Math.* 288 (1966):239-247.
2. H. Halberstam & K. Roth. *Sequences* (Chapter 4). Oxford: Oxford University Press, 1966.
3. *MECC Elementary Volume 1 Mathematics (Games and Drills)*. Minnesota Educational Computing Consortium, 3490 Lexington Ave. North, St. Paul MN 55112, 1983, pp. 47-52.

◆◇◆◇◆