Isabelle Sanford
isanford@brynmawr.edu

The Taxman

What We Know

ABSTRACT. I describe the taxman game and discuss three winning strategies, as well as some bounds on the best possible score.

1. INTRODUCTION

The taxman game is a computer game and introductory programming exercise where the player picks a number each turn from a pot of the first n natural numbers. The "taxman" takes every proper divisor of the player's pick, and the taxman must take something each turn. The game ends when the player can pick no further numbers (i.e. there are no remaining numbers with proper divisors that haven't been taken by the player or the taxman), at which point the taxman takes anything unpicked. The goal, naturally, is to score more points than the taxman. You can try the game yourself, **here**.

A sample game might play as follows: first, pick a pot size, say $n = 20$. On the first turn, the player can pick any natural number up to 20, except for 1 (which has no proper divisor the taxman could take). Let's try the most obvious strategy: just picking the highest available number each time, starting with 20. Now the player has {20} but the taxman gets all its proper divisors: {1, 2, 4, 5, 10}. So we're already behind, with a score of 20 to the taxman's 22. Going down, 19 is not available (its only proper divisor is 1, which the taxman has already taken), but we can pick 18. Then the taxman gets 3, 6, and 9, and now has {1, 2, 3, 4, 5, 6, 9, 10} = 40 to our 38. 17 isn't available, but 16 is, and the taxman only gets an additional 8 from that (having already taken all the other proper divisors of 16), so we're finally ahead with 54 to the taxman's 46. Then we take 14, the taxman gets 7, and... there are no remaining moves. We have {14, 16, 18, 20} and the taxman has {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, which would be great if the taxman did not also now get all of the remaining unpicked numbers, {11, 12, 13, 15, 17, 19}, putting us at a sad 68 to the taxman's 142.

This paper will detail two strategies (described in Moniot (2007) and Perlmutter (2015)) that work better than this one in section 2 and then discuss the Hensley (1985) paper that examines the best possible score in section 3.

2. WINNING STRATEGIES

Perlmutter and Moniot both suggest possible winning strategies, detailed below in subsection 2.2 and subsection 2.3. Moniot's strategy wins for at least all $n < 1000$ and Perlmutter's is proven to win for an unspecified 'sufficiently large' n. Hensley also suggests a strategy as part of his discussion of the optimal score, which will be laid out at the end of section 3.

2.1. **The First Pick.** While multiple winning strategies are laid out in this paper, we know very little about the *optimal* strategy, except with regards to the very first pick. Because 1 is a proper divisor of every integer greater than 1, the taxman will always take 1 on the first pick. That in turn will mean that on future turns the player will no longer be able to pick any number with only a proper divisor of 1, i.e. all of the prime numbers. Thus we will want to pick the largest prime available, and the taxman will (eventually) take the rest.

2.2. **Moniot's "improved greedy heuristic".** The most naive strategy possible, as we tried in section 1, would be to simply pick the largest number each time. We saw that this won't always win (and in fact will usually lose), and its major flaw is that it ignores the taxman's take each turn. Moniot proposes instead a variation on this idea in which you pick the number with the largest *net* gain for you (that is, the number which maximizes the difference between your score and the taxman's). This generally does well, but in trying the strategy Moniot noticed that its next recommendation sometimes neglected

a 'freebie' number, a proper divisor of the recommendation that could be chosen first without removing the recommendation or otherwise affecting the rest of play. Scanning for these 'freebies' between each pick and otherwise maximizing the net gain each turn is the best strategy Moniot found, gives the player about 60% of the pot on average in the range up to $n = 1000$, and always wins in that range. It is not, however, always the optimal strategy (Moniot computed the optimal sequence up to $n = 49$ and found that his strategy was optimal about half the time), and he conjectures but has no proof that it will always be a winning strategy.

2.3. **Perlmutter's "asymptotically winning strategy".** In many ways, Perlmutter's strategy is the opposite of Moniot's. He does provide a proof that it is a winning strategy for sufficiently large n (but does not know what that n is), but can only conjecture that it wins for $n > 19$. The proof is not detailed here, but the strategy is as follows.

For every number k from 0 up to half of the pot size $(n/2)$, let a_k be the largest multiple of k in the upper half of the pot with no proper divisors greater than k. If no such multiple exists, then a_k is undefined. The strategy is simply: pick each (defined) a_k in order starting from $k = 1$. After every number in the sequence has been picked picked, the player has enough points to win the game, so if the game is not yet over pick any legal moves until it is.

3. The Optimal Score

Let $f(n)$ denote the best possible score for a pot size of n. The total number of points at stake in a game is the sum of the natural numbers up to n, i.e. $n(n + 1)/2$. Moniot points out that we know the taxman and the player must both take at least one number on each turn, so the best possible score for the player would occur if the player took every number in the upper half of the pot and the taxman took every number in the lower half. That means the taxman takes 1 through $n/2$, which sums to $n(n + 2)/8$ points. Dividing this by our total pot size, we get a taxman's take of $\frac{1}{4} \cdot (n + 2)/(n + 1)$, which is always a little bit bigger than one quarter. So at best the player can get a little less than 75% of the pot[1].

When we consider n for very large numbers, we can approximate our pot size to $n^2/2$. Let C be the fraction of the pot that the player can win, that is,

$$C = \frac{f(n)}{\frac{1}{2}n^2}$$

for sufficiently large n. There is no proof for a particular strategy that will always achieve the best score, but Hensley does prove the following:

(1) The lower bound of C exists and is at least $1/2$, which means that $f(n)$ is always a winning strategy for sufficiently large n.
(2) The upper bound of C exists and is at most $35/48$.
(3) C converges to a limit as n goes to infinity.

He also outlines a strategy that he claims (without proof) will win for pot sizes larger than six million.

This section will explain the proofs for the existence of C's upper and lower bound, though not their convergence, and then detail Hensley's winning strategy and his numerical bounds on C.

3.1. **Definitions.** Hensley's primary strategy is to partition the pot into subsets which can be played individually in smaller and more manageable games. He easily uses these partitions to show a lower bound (basically the sum of the best scores on each individual partition), and then he examines a hypothetical sequence of moves and uses the same partitions to find a rough upper bound.

---

[1]The player can in fact take all of the upper numbers for $n = 10$, and ends up with $40/55 = 73\%$ of the pot, which may well be the best possible score for any n.

First, pick[2] a prime p in your pot. Let $\mathbf{D_p}$ be the set of natural numbers whose prime factors are all p or less. For example, D_3 contains only the natural numbers whose prime factors are $2^n 3^m$ for some integer $n, m \geq 0$, that is, $D_3 = \{1, 2, 3, 4, 6, 8, 9, 12, 16, \dots\}$. Let $\mathbf{A_p}$ be the product of all primes up to p, e.g. $A_3 = 2 \cdot 3$. The symbolic definitions of D_p and A_p that Hensley provides are given below for completeness, but are more likely to confuse than to clarify.

$$D_p = \{d : (q|d \wedge q \text{ prime}) \to q \leq p\} \qquad A_p = \prod_{q \leq p, \ q \text{ prime}} q$$

Let us also define $\mathbf{f_p(x)}$ as the best possible score for the player playing on the (non-standard) initial set of D_p from 1 up to x. (Here x can be any real positive number, but e.g. $f_3(4)$ and $f_3(5.5)$ are the same score because the next number in D_3 is 6.)

Now we can describe the partitions of our pot $\{1, 2, \dots, n\}$. There is one partition for every k relatively prime to A_p, that is, any number whose prime factors are all greater than p (plus a partition for $k = 1$). Let the set $\mathbf{N_{k,p}(n)}$ be defined as

$$N_{k,p}(n) = \{kd : d \in D_p \wedge kd \leq n\}$$

That is, for some prime p on a pot size of n, each partition is D_p multiplied by a particular k relatively prime to A_p, with any element that ends up greater than n removed. In practice, what this means is that $N_{1,p}(n)$ will be the subset of D_p which is smaller than our pot size n. Every subsequent $N_{k,p}(n)$ will be the result of multiplying each element in D_p by k, and taking the subset of that up to n. So for instance, for $p = 3$, the partitions will be

$$N_{1,3} = \{1, 2, 3, 4, 6, 8, 9, 12, 16, 18\},$$
$$N_{5,3} = \{5, 10, 15, 20\}, N_{7,3} = \{7, 14\},$$
$$N_{11,3} = \{11\}, \ N_{13,3} = \{13\}, \ N_{17,3} = \{17\}, \text{ and } N_{19,3} = \{19\}.$$

3.1.1. *$N_{k,p}$ is a partition.* Let x be an arbitrary element of our pot, and imagine dividing its prime factors into two groups, the primes greater than p and those less than or equal to p. The product of the primes greater than p will be relatively prime to A_p, which means it will be the k for one of our partitions. The product of primes in x up to p will be an element of D_p, since D_p contains every element whose prime factors are only primes less than or equal to p. Thus any x can be represented as kd for $d \in D_p$ and k coprime to A_p. And we know that kd will be an element of $N_{k,p}$, so every x in the pot is definitely part of at least one $N_{k,p}$.

To be a partition, we must also show that the $N_{k,p}$ are disjoint. Suppose on the contrary that there exists some x which is an element of $N_{k,p}$ and $N_{k',p}$ for $k \neq k'$. As before, we can break down x's prime factors into the primes up to p and the primes above p, and represent those as d and k. So $x = kd$, and $x = k'd'$, with $k \neq k'$. This implies that the prime factors of x have been divided up differently between kd and $k'd'$. But neither k nor k' can have a prime factor less than or equal to p (or they would not be coprime to A_p). Removing a prime factor from k' and putting it in d' would result in d' no longer being an element of D_p, and then $k'd'$ would no longer be an element of $N_{k',p}$. So the factors of k and k' must be exactly the same, which is a contradiction. Thus every element in the pot must be a member of exactly one $N_{k,p}$, and $N_{k,p}$ is a partition on $\{1, 2, \dots, n\}$.

3.1.2. *Playing on a partition $N_{k,p}$ will not disturb partitions with higher k's.* Consider, for instance, $N_{1,3}(40)$ and $N_{5,3}(40)$:

$N_{1,3}$	1	2	3	4	6	8	...
$N_{5,3}$	5	10	15	20	30	40	

Suppose we play on $N_{1,3}$ such that every single element of it is taken (by either the player or the taxman), and then move to $N_{5,3}$. We know that 5 is no longer available to be picked (it is prime, and 1 is taken), but it isn't actually taken yet, which means that every multiple of 5 is still available at the moment, i.e. the rest of $N_{5,3}$ must be able to be picked.

In more general terms, if we play on $N_{k,p}$ and then examine $N_{k',p}$ for $k' > k$, each element in the latter can be represented as $k'd$ for some $d \in D_p$. We know k' cannot yet have been taken (the player cannot have taken it because $k' \in N_{k',p}$ and partitions are disjoint; the taxman cannot have taken it because it is not a proper divisor of any smaller k), so k' is an unchosen proper divisor for all $k'd$, and thus every element of $N_{k',p}$ (except for k' itself) must still be available to be picked.

3.2. **A Lower Bound for** C. First, let us examine the best possible score on a particular partition $N_{k,p}(n)$. Since each element is k times the corresponding element in D_p (that is, the xth number in the partition will be k times the xth number in D_p), we can equivalently call this the best possible score on D_p (up to some number), times k. That 'some number' is the largest $d \le n/k$ (because the largest number in the partition is $kd \le n$), so we can write the best score for a partition as $k \cdot f_p(n/k)$.

Given that we can play the partitions in increasing order without disruption, we can sum over the score for each partition and get

$$\sum_{k \le n}^{*} k f_p(n/k)$$

where the $\sum^{*}$ indicates that our sum is only over the relevant k (that is, k coprime to A_p). This score is a lower bound for $f(n)$:

(1) $$f(n) \ge \sum_{k \le n}^{*} k f_p(n/k)$$

That is, if we know how to play optimally on D_p, we can play on each partition in turn, and the best possible score must be at least as good as that result.

Next, we can notice that there are usually some partitions with the same number of elements. Say we examine the partitions for $p = 3$ and $n = 40$ and see that both $N_{11,3}(40) = \{11, 22, 33\}$ and $N_{13,3}(40) = \{13, 26, 39\}$ have three elements. This is helpful because we are playing the exact same set of moves for each: that is, any $N_{k,p}$ can be played on the subset of D_p with the same number of elements and the score is just multiplied by k at the end. Since $N_{11,3}(40)$ and $N_{13,3}(40)$ are using the same subset of D_3 (i.e. $\{1, 2, 3\}$), their best scores would be $11 f_3(3)$ and $13 f_3(3)$, respectively. So we could say that the sum of the best scores of all the length-3 partitions is $f_p(3) \sum^{*} k$, where we know that those k will be between $n/3$ and $n/4$ (since 4 is the next element in D_p, so any k smaller than $n/4$ will have at least 4 elements, not just 3). Combining this with the sum from above in (1), we can take the sum of the best scores for each subset of partitions with j elements, and have

(2) $$f(n) \ge \sum_{j \in D_p}^{n} f_p(j) \sum_{\frac{n}{j'} < k \le \frac{n}{j}}^{*} k$$

where j' indicates the next element in D_p.[3]

---

[3]On notation: Hensley uses $\sum_{j \in D_p, j \le n}$ for the outer sum here, but having n at the top seems both cleaner and more logical, and I can see no loss of information (but could be wrong, thus this note).

Now, suppose we define[4] B_p as some integer b divided by A_p, such that for any A_p consecutive numbers, b of them are relatively prime to A_p. For instance, for A_3 we know that of any 6 consecutive numbers, three will be even (i.e. divisible by 2), two will be divisible by 3, and those two groups will overlap in the number divisible by 6. So in total there will be 2 remaining numbers that are relatively prime to 6, meaning that in this case $B_p = 2/6 = 1/3$. B_p effectively represents the fraction of numbers in a consecutive sequence that will be coprime to A_p.

We can use this to simplify the inner sum in (2), since we're summing over some consecutive sequence of numbers ($\frac{n}{j'}$ to $\frac{n}{j}$) for only those numbers coprime to A_p. Then the approximate number of terms in our sum will be $B_p(\frac{n}{j} - \frac{n}{j'})$. The average value of a term in that sum will be the mean of our consecutive sequence, i.e. $\frac{1}{2}(\frac{n}{j} + \frac{n}{j'})$. So multiplying this average by the number of terms implies

$$(3) \qquad \sum_{\frac{n}{j'} < k \le \frac{n}{j}}^{*} k \approx \frac{1}{2} n^2 B_p \left(\frac{1}{j^2} - \frac{1}{j'^2} \right)$$

Then substituting this into (2), we get

$$(4) \qquad f(n) \ge \frac{1}{2} n^2 B_p \sum_{j \in D_p}^{n} f_p(j) \left(\frac{1}{j^2} - \frac{1}{j'^2} \right)$$

This is our lower bound on $f(n)$. (We use this form because we can make our upper bound come out to a very similar form, and then the limit if the bounds converge is easy to define.)

3.3. **An Upper Bound for C.**

To determine an upper bound, let's first suppose we choose[5] some prime p and then play a game on $\{1, 2, \ldots, n\}$. Each turn, we pick some number m, and we note down the largest divisor in play at the time (i.e. the largest number the taxman takes from our pick), and call that number $t(m)$. Each $t(m)$ is definitely distinct because the taxman does not take numbers twice; once an element is taken it is no longer a possible future $t(m)$. Then we separate the m we picked into two sets: M_1, for which the greatest available proper divisor $t(m)$ was smaller than n/p, and M_2 where $t(m) \ge n/p$. That is:

$$M_1 = \{m : t(m) < n/p\} \qquad M_2 = \{m : t(m) \ge n/p\}$$

There can only be n/p elements in M_1, since each must have a distinct picked divisor in the set $\{1, 2, \ldots, n/p\}$. So the score from M_1 must be smaller than n^2/p, i.e. n/p elements which can each be worth up to n (only one element is actually worth n, but since we're looking for an upper bound an overestimate is fine here). This is an upper bound on M_1, but a useful[6] upper bound on M_2 is trickier.

Hensley then uses a lemma. In his words:

> Lemma 1: Suppose k is relatively prime to A_p, d ∈ D_p, and kd ∈ M_2. Then k|t(kd).

That is, suppose we pick some m that is in M_2. Just like in 3.1.1, we can represent that m as kd, where k is the product of prime factors greater than p and d is the prime factors less than or equal to p. The largest unchosen proper divisor of m, $t(m)$, will be kd divided by some number a. If $a \in D_p$, all its prime factors come out of d, and $t(m) = kd/a$ will still be divisible by k as required. If one of the prime factors of a is larger than p, then kd/a must be smaller than n/p (since $kd \le n$ and $a > p$) and thus cannot be in M_2.

---

[4]Hensley initially defines $B_p = \prod_{q \le p}(1 - 1/q)$ for only prime q, but this is the only context in which he uses it. For instance, $B_3 = (1 - 1/2)(1 - 1/3) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$. We can see that this will end up as some integer divided by A_p, because the denominator of each term is q, and A_p is the product of those q's.

[5]Hensley again makes no comment on how to choose a prime, but here it will mostly just serve to divide our pot into two groups.

[6]If we followed the same logic for M_2 as M_1, we'd be back to Moniot's logic in the first paragraph of section 3 and find an upper bound of about 3/4. But since we know that $t(m)$ is always strictly less than our pick m, we can get a narrower upper bound.

Now for each k (i.e. each number less than n and relatively prime to A_p), consider the subset of M_2 that can be represented as kd for some $d \in D_p$. We can imagine this subset as one of our partitions $N_{k,p}(n)$, and thus could play a game on D_p up to n/k. A choice of a d in D_p would be equivalent to a choice of kd in M_2. If each choice of d does in fact correspond to a legal move in M_2 (which we will show in a moment), then we can say that the best possible score of this subset will be $kf_p(n/k)$. (The best score on D_p up to n/k is $f_p(n/k)$, and each choice in the 'real' game is the element of D_p times k.) That is,

$$(5) \qquad \sum_{d\in D_p,\ kd\in M_2} kd \leq kf_p(n/k)$$

Each d is a legal move. Since $t(kd)$ is a proper divisor of kd, we can equivalently say that $\frac{t(kd)}{k}$ is a proper divisor of d (if $t(kd)/k$ is an integer, which it must be from Lemma 1). We also know that $t(kd)$ had not already been deleted before kd was chosen, which means that no multiple of $t(kd)$ has been chosen. We can represent such a multiple as kd' (with similar reasoning to that used in lemma 1: multiplying $t(kd)$ by a factor of k would result in a number larger than our pot size, so every possible pick that's a multiple of $t(kd)$ must be $t(kd)$ times some element of d). And if no kd' has been picked in the real game, then no multiple d' of $t(kd)/k$ has yet been chosen in the equivalent D_p game. So $t(kd)/k$ must still be available in the D_p game. $\square$

Summing each side of (5) over k gives the sum of all elements in M_2, i.e. $\sum_{k\leq n}^* f_p(n/k)$. Adding this to our previously found best score for M_1 results in

$$\sum_{m\in M_1\cup M_2} m \leq n^2/p + \sum_{k\leq n}^* f_p(n/k)$$

and we can group k's just as we did with the lower bound to get

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

which is an upper bound for $f(n)$. We can then use (3) again to replace our inner sum, which (pulling out $\frac{1}{2}n^2$) gives us a final upper bound of:

$$(7) \qquad f(n) \leq \frac{1}{2}n^2\left(\frac{2}{p} + B_p\sum_{j\in D_p}^n f_p(j)(\frac{1}{j^2} - \frac{1}{j'^2})\right)$$

3.4. **Convergence.** To review, we have the following lower and upper bounds:

$$f(n) \geq \frac{1}{2}n^2 B_p\sum_{j\in D_p}^n f_p(j)(\frac{1}{j^2} - \frac{1}{j'^2})$$

$$f(n) \leq \frac{1}{2}n^2\left(\frac{2}{p} + B_p\sum_{j\in D_p}^n f_p(j)(\frac{1}{j^2} - \frac{1}{j'^2})\right)$$

Also recall that C is the best possible score divided by the number of points in the pot, which approximates to $\frac{1}{2}n^2$ for large n. So we can combine the above two equations, assuming they converge, and get

$$(8) \qquad C = \lim_{n\to\infty}\frac{f(n)}{(1/2)n^2} = \lim_{p\to\infty} B_p\sum_{j\in D_p} f_p(j)\left(\frac{1}{j^2} - \frac{1}{j'^2}\right)$$

Hensley's proof that (4) and (7) do in fact converge is omitted here for brevity. (He does not prove a particular numerical value for C.)

3.5. **Concrete Bounds and a Winning Strategy.** Hensley then shows using the bounds explained above that $0.5 < C < .75$, and states that in fact $.56 < C < .729$. He also outlines the following strategy (which he claims, but does not appear to prove, will win for $n > 6,000,000$):

> (A) Partition $\{1, 2, \ldots, n\}$ into sets of the form $N_{k,5}(n) = \{kd : d \in D_5, d \leq n/k\}$ 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 $(n/200) \leq k \leq n$, play $N_{k,5}$ as instructed by the table[7]. Start with smaller values of k and work up.

4. Conclusions and Open Questions

This paper laid out two strategies for the Taxman game, each of which win under a different set of conditions, and then examined Hensley's paper in regards to what we know about the best possible score for a game as n gets very large. We set out the full proofs that that score has upper and lower bounds, omitting the proof that they converge, and finally outlined Hensley's claimed winning strategy for $n > 6,000,000$.

To the best of my knowledge, this list is all we know about strategy and score for the Taxman game:

- The optimal first pick is always the largest prime. (Moniot)
- There is a known strategy that will win every time for sufficiently large n. (Perlmutter)
- The fraction of the pot that the player can win converges to a limit, somewhere between .56 and .729. (Hensley)
- The player can never win more than 75% of the pot. (Moniot)

Moniot claims on his research page that the game is probably NP-hard. The optimal sequence appears to have been found for each pot up to $n = 158$, here, and the best possible score is available up to $n = 719$ here.

With such a short list of knowledge, many open questions remain. A few are listed here:

- Aside from $n = 8$ and $n = 20$, is the optimal second pick always the largest square of a prime? (from Moniot)
- If the Taxman were played as a 2-player game (i.e. each player takes turns picking, and the other player get's the taxman's take that turn), do any of these winning strategies apply? Can the first player always win, and by how much? (from Perlmutter)
- Do the Moniot and Perlmutter strategies both always win for $n > 3$ and $n > 19$ respectively?
- Is there a simple formula for the optimal strategy for any n?
- Do all three of the given strategies always make the optimal first pick (i.e. the largest prime)? (I suspect the proof that Perlmutter's strategy does would be fairly simple.)
- Can a specific n be proven for Perlmutter's 'sufficiently large n'? Can Hensley's $n > 6,000,000$ be reduced to something more manageable?

5. References

HENSLEY, DOUGLAS, *A Winning Strategy at Taxman*, The Fibonacci Quarterly, Vol. 26, No. 3, pp.262-270, 1988.

MONIOT, ROBERT, *The Taxman Game*, Math Horizons, pp.18-20, Feb. 2007.

PERLMUTTER, NORMAN LEWIS, *Beating the Taxman Asymptotically*, Pi Mu Epsilon Journal, Vol. 14, No. 3 (Fall 2015), pp.199-204, 2015. (Revised version of paper first published in 2007.)

---

[7]This table is provided in the paper, and is a list of moves on subsets of D_5 from 1 to 36 elements.

6. Further Reading

(not yet incorporated into this paper)

Wilson, Brandee, *The Taxman Game*, 2009. Available online: https://web.archive.org/web/20151023160904/http://scimath.unl.edu/MIM/files/MATExamFiles/Wilson_FINAL.pdf. Accessed December 2021.

Moniot, Robert, *The Taxman Game*, somewhat expanded from the version in Math Horizons. Available online: https://www.dsm.fordham.edu/~moniot/taxman-game.pdf. Accessed December 2021.

These and other resources are available at the Online Encyclopedia of Integer Sequences (OEIS) entry for the sequence that represents the best scores in the Taxman game for each pot, here: https://oeis.org/A019312.