"If you ponder your moves well, it is possible to beat the taxman."

The Taxman Game

Robert K. Moniot
Fordham University

Want to know how to beat the taxman? Legally, that is? Read on, and we will explore this cute little mathematical game.

The taxman game is a "golden oldie" computer game. It has been used as an exercise in introductory programming classes and in books on recreational computing since time immemorial. (Well, at least since the 1970s. I haven't been able to determine who invented it or when.)

Here's how to play: you start with a pot consisting of all the positive integers up to some chosen limit, *N*. You take one, and the taxman gets all the others that divide it evenly. The number you picked is added to your score, and the numbers chosen by the taxman are added to his score. This process is repeated until there's nothing left in the pot. What makes the game interesting is that the taxman has to get something on every turn, so you can't pick a number that has no divisors left. And when the numbers remaining in the pot have no divisors left, the taxman gets them all!

The game is pretty easy to program on a computer, and if programming is one of your skills, you might want to pause now to do so and play a few games, just to get the hang of it. Right now we'll demo a couple of example games with $N = 10$ to show you how it goes. First, a game where we try to win by always taking the largest number in the pot.

```
Hi! I'm the taxman! Let's start.
How many numbers in the pot? 10
The available numbers are: 1 2 3 4 5 6 7 8 9 10
Pick one of the numbers: 10
I take: 1 2 5
The available numbers are: 3 4 6 7 8 9
Pick one of the numbers: 9
I take: 3
The available numbers are: 4 6 7 8
Pick one of the numbers: 8
I take: 4
I take: 6 7
All gone. My score = 28, your score = 27
I win.
```

As you can see, it's not so easy to beat the taxman. You do have an advantage because each number you take is always bigger than any of the numbers the taxman takes, and so it's not hard to get out in front of the taxman and stay there while play continues. But slowly and silently the unpickable numbers with no divisors left are accumulating, and suddenly the game ends with the taxman's score getting a big boost. If you haven't accumulated enough of a cushion, he'll jump ahead and beat you. In this game, the numbers 6 and 7 ended up with no factors left, so the taxman got them, boosting his score from 15 to 28.

Still, if you ponder your moves well, it is possible to beat the taxman. This time we'll choose more carefully.

```
Hi! I'm the taxman! Let's start.
How many numbers in the pot? 10
The available numbers are: 1 2 3 4 5 6 7 8 9 10
Pick one of the numbers: 7
I take: 1
The available numbers are: 2 3 4 5 6 8 9 10
Pick one of the numbers: 9
I take: 3
The available numbers are: 2 4 5 6 8 10
Pick one of the numbers: 6
I take: 2
The available numbers are: 4 5 8 10
Pick one of the numbers: 8
I take: 4
The available numbers are: 5 10
Pick one of the numbers: 10
I take: 5
All gone. My score = 15, your score = 40
You win.
```

In this game, we limited the taxman's take to just one number on each move, and no numbers were left unpickable at the end. It's not always possible to trounce the taxman this badly, but it is generally possible to win. Below, we will explore some strategies for making good picks. We will also see the results of a lengthy computer search for the best possible sequences of picks for pots of up to 49 numbers.

Some basics of strategy

First of all, it is worthwhile noticing that since each of the *N* numbers originally in the pot ends up in either the player's score or the taxman's score, the sum of the two scores must equal $N(N + 1)/2$, the sum of the numbers from 1 to N. Therefore the bigger the taxman's score, the smaller your score, and vice-versa. This kind of game is called a "zero-sum game," because the sum of the two players' final scores, minus a constant that is known at the start of play, is equal to zero.

The taxman always gets the number 1 and all but at most one of the primes that are initially in the pot. This is because the number 1 can never be picked, and the only turn on which the player can pick a prime number is the first turn, when the number 1 is still in the pot. This gives the taxman a built-in advantage. There is no simple formula for this advantage, but a crude estimate can be calculated by using the approximation for the density of primes near x of $1/\ln x$. We will skip the details of the calculation, but the result is that for a pot of size N, the taxman is guaranteed to get roughly $1/\ln N$ of the pot just from the leftover primes. This fraction slowly decreases as N increases: it is about 30% for $N = 30$, and about 15% for $N=1000$.

On the other hand, we can ask what is the absolute best that the player can do in any game? Since the taxman must get something on every turn, the best you can hope for is to take just as many numbers as the taxman does. This can only be done for even N. In that case, the best score you can get is if you take all of the largest numbers and leave the smallest ones for the taxman. (We were in fact able to do this in the second example game played above, with $N = 10$.) So for even N, if the player takes all the numbers from $N/2 + 1$ through N, leaving the taxman the numbers 1 through $N/2$, a little algebra shows that the taxman's score will be $N(N + 2)/8$. Since the pot is $N(N + 1)/2$, the taxman's take as a fraction of the pot is $1/4$ $(N + 2) / (N + 1)$. This fraction is always a bit bigger than 1/4. This means that the very best the player can ever do is to get a bit less than 75% of the pot.

Optimal strategy

An *optimal strategy* for this game would be a strategy that always gives the player the largest possible score for a given N (and gives the taxman the least). What can we say about the optimal strategy?

The best choice for the first pick is not hard to see: simply take the largest *prime* number in the pot. We can prove that this is always the best pick as follows. First, this pick limits the taxman's take to 1, which the taxman will take on the first move for any pick. Further, removing this prime, which we will call $p_{\max}$, and 1 from the pot does not prevent any sequence of nonprime picks that was originally possible. That is, if the player could have played by picking the sequence of numbers $n_1, n_2, \ldots, n_k$ where k is the total number of picks and n_1 is not prime, then the alternative sequence $p_{\max}, n_1, n_2, \ldots, n_k$ is also possible and gives the player a higher score. Picking $p_{\max}$ does not remove any divisors of $n_1, n_2, \ldots, n_k$ other than 1, because $p_{\max}$ cannot divide any other number in the pot. Its smallest multiple, $2p_{\max}$, cannot be in the pot, because there is a theorem from number theory (postulated by Bertrand and proved by Chebyshev) that says that there is always a prime between $p_{\max}$ and $2p_{\max}$, which would contradict the choice of $p_{\max}$ as the largest prime in the pot. Since none of the picks in the original sequence is prime, they must therefore all have other divisors in the pot that will still allow them to be picked.

On the other hand, if n_1 is prime and $n_1 < p_{\max}$, then by a similar argument replacing n_1 by $p_{\max}$ does not require any change in the rest of the sequence, and increases the player's score by $p_{\max} - n_1$.

From this result, it follows that if N itself is prime, then the optimal sequence of picks for a pot of size N is the same as for a pot of size $N - 1$, except that on the first turn we should pick N instead of the next-smaller prime.

A heuristic strategy

Beyond the first pick, there don't seem to be any simple rules for the optimal strategy. In situations like this, we can often be happy to settle for a strategy that may not be optimal but will generally give a high score. Such a strategy, which is not proved to be the best but is sufficiently successful in practice, is called a *heuristic*. Heuristics can be based on various principles. For the taxman game, a "greedy" principle yields a very good heuristic that is easy to apply. A greedy principle is one that says to make the move that maximizes the short-term gain. For some problems, a greedy strategy can turn out to be optimal, but in many cases it falls short because of the long-term consequences. (The same is true of life in general, isn't it?)

In the first example game we saw a naive greedy heuristic that failed even to win the game: always picking the largest available number in the pot. This is a bad heuristic because it ignores the taxman's take. The game is zero-sum, so the true measure of gain is not your take alone, but the difference between your take (the value of the number you pick) and the taxman's take (the sum of the remaining divisors of the number).

Therefore, a better greedy heuristic for the taxman game is on each turn to pick the number that maximizes the difference between your pick and the taxman's take for that pick. This heuristic is easy to calculate (at least, easy for a computer) and generally gives a good score.

Trials of the greedy method show that it sometimes overlooks "freebies," which are picks that can be taken in between greedy picks without altering the remaining sequence. The freebies are numbers for which the taxman's take is a proper subset of his take for the greedy pick on the same turn. That is, if you picked the freebie, the taxman would take some, but not all, of the divisors of the greedy pick, so the greedy pick could still be taken afterward. Thus the greedy pick would either give the freebie to the taxman or make it unpickable subsequently, whereas picking the freebie before the greedy pick does not prevent the greedy pick. (The freebie also must not be a divisor of any other number in the greedy sequence, lest it make that number unpickable.)

For example, when $N = 15$, after first picking 13, the largest prime, the greedy pick is 15. The remaining divisors of 15 are 3 and 5. This means that we can pick 9, giving the taxman 3, and still pick 15, giving the taxman 5. This makes 9 a freebie.

Therefore the greedy method can be improved by scanning for freebies before each greedy pick, and inserting them into the sequence. I call this the "improved greedy" heuristic, and it is the best strategy that I have found so far. But is it optimal? Alas, only sometimes. The experiments described later show that the improved greedy heuristic often falls short of the optimal score, but it does beat the taxman for all N up to 1000.

Results of the search for optimal picks

Since there does not seem to be any simple strategy that guarantees an optimal score, I used a computer to search for the optimal play for any given N. This search was carried out for all pots up to $N = 49$ before I reached the limit of my computing resources. The results are available online at `www.maa.org/MathHorizons`.

One interesting question that we can consider using these results is: how often can we achieve the absolute best score, in which we take all the numbers in the upper half of the pot? The search shows that this is possible only for the cases $N = 2, 4, 6,$ and 10. For $N = 8$, and for all N above 10, there are at least two primes in the upper half of the pot, so the taxman always gets at least one of them. For $N = 10$, the player's score is $40/55 \approx 73\%$ of the pot. This is close to the upper limit of 75% derived previously. It is probably the best score, as a fraction of the pot, that can be achieved for any pot size.

Another question is: how well does the improved greedy heuristic do as compared to the optimal play? The results show that the heuristic fails to yield the optimal play for about half the cases solved. Still, it always does a good job. Since the improved greedy heuristic can be calculated quickly, I explored how well it does out to $N = 1000$. In this range, it always makes the optimal first pick and it always wins (except for the case $N = 3$ in which the optimal score is a tie). On average, it gives the player 60% of the pot.

The experimental results exhibit an interesting feature: the optimal *second* pick is almost always the largest *square* of a prime. The only cases I found where this pick is not optimal are $N = 8$ and $N = 20$. (Some of the sequences could be rearranged to pick it later than second and still get the same score.)

Open questions

There are many questions about the taxman game that I haven't been able to answer. Here are a few.

Does the greedy heuristic always make the optimal first pick, namely the largest prime? It seems likely that it does, since the largest prime is usually one of the biggest numbers in the pot, and so the difference between it and 1 (the taxman's take for that pick), is going to be hard to beat. I have tested it out to $N = 500{,}000$ without finding any cases where it fails to make the correct first pick. Still, I don't have a downright proof.

Is there a simple *winning strategy* for the game? This would be a strategy that is guaranteed to win every time, even if it's not optimal. The improved greedy heuristic is a good candidate, and tests show that it wins all games up to $N = 1000$, but I have not proved that it always wins.

Is the choice of the largest squared prime as the second pick optimal for all N above 20? If not, how can we identify the values of N for which it is not?

We saw how the optimal play for prime N can be based on the optimal play for $N–1$. Are there any other cases for which the optimal play for one pot size can be used to find the optimal play for another? The experiments show that very often the optimal sequences for successive N are quite similar, but there isn't an obvious systematic way to relate them.