THE MOST IMPORTANT ANSWER TO THE MOST IMPORTANT QUESTION

K.M. HALPERN

Abstract. Should a husband leave the toilet seat up or down? It turns out that deriving an optimal strategy which burdens both partners equally is nontrivial. We seek an optimal history-independent fair strategy using a sensible cost function. The limited set of deterministic strategies proves inadequate except in contrived instances, so we derive an optimal fair randomized strategy. To do so, we construct an appropriate Markov chain, identify its stationary state, and find a choice of strategy parameters which minimizes the cost function in that state subject to the fairness constraint.

1. Introduction

There is a question which has plagued humanity since the beginning of time, the subject of endless philosophical debate. This is whether to leave the toilet seat up or down. Unfortunately, since all great thought takes place on the toilet, there could be no philosophy prior to its invention by Tom Bradney in 1927. This is why Rodin's famous sculpture shows a man, head in hand, sitting on a rock. He clearly had assumed the position and wanted to think, but could not. It did not help that he was made of bronze. Though he could not comprehend what was missing from life, after 1927 we found out: marital strife born of a mispositioned toilet seat. We now offer the next stage in human evolution: a mathematical solution to that conflict. As any unmarried scientist knows, such a mathematical solution will find ready acceptance in every household.

We begin by putting aside all notions of chivalry, because chivalry has no place in a modern enlightened society. We also put aside questions of marital (or cohabitative) tranquility, because marriage and tranquility have no place in a modern enlightened society. Finally, we completely ignore any aesthetic considerations, because aesthetic considerations have no place in a modern enlightened society. And while the sensible amongst us avoid the rigors of a shared bathroom, the present exercise nonetheless has the substantial intellectual merit of earning the author a publication credit. More are necessary, so let us attend to this matter with alacrity.

We refer to the genders as M and F. For those who identify as both, neither, or some linear combination of the two (perhaps time-varying), we note that the only salient consideration for our present calculation is anatomical. M herein refers to any being which urinates standing up and defecates sitting, while F refers to any being which performs both functions sitting. Couples or roommates of the same anatomical gender have a simple, obvious solution. In

Date: June 8, 2022.

2010 *Mathematics Subject Classification.* Research paper.

Key words and phrases. behavioral economics, markov chains, markov decision processes, game theory, fairness.

that case, any toilet seat problems are of their own invention and beyond the scope of our work. Nor does out analysis apply to those who do not fit into the M or F categories due to inclination, anatomy, or political stance. Such individuals are encouraged to employ the methods described herein to calculate solutions better suited to their particular situation. They also are encouraged to cite this article when doing so, preferably multiple times, because citations count toward tenure, tenure means money toward toilet seat upgrades, and better toilet seats mean better thoughts, thus advancing humanity.

The two relevant bodily functions are affectionately referred to herein by their popular designations: #1 and #2. We do not consider other possible functions $\#3 \dots n$, as these unnecessarily complicate an already unnecessary calculation. Presumably, they are of infrequent occurrence. If not, the problem is bigger than one of toilet seat position. A doctor, psychiatrist, or mechanic may prove of greater benefit than the optimization of an artificial utility function.

2. Framework

The layout of our problem is simple. Two participants, one M and one F, each use the same bathroom. Sometimes they need to perform #1 and sometimes #2. The toilet seat can have two positions, which we'll denote U and D. The F user requires D for both operations, while the M user requires U for #1 and D for #2.

2.1. Transaction.

For a given instance of use, one of the two participants (M or F) enters the restroom to perform a specific function (#1 or #2). They find the toilet seat in the position (U or D) it was left after the previous use.

Before use, the toilet seat must be in the correct position for the participant and function. As a result, the user may need to raise or lower it.

After use, the participant may decide to leave the toilet seat in a particular position. This may entail raising or lowering it.

We'll refer to this whole process as a Transaction. For our purposes, life can be considered an endless sequence of such transactions. This makes it marginally worse than perdition and marginally better than having a roommate.

2.2. Basic Assumptions.

We make some core assumptions regarding each transaction: (i) the choice of user M or F is random, IID with equal (0.5) probability and (ii) the choice of bodily function is random, IID with some fixed, known probability p of being #1. Transactions are independent from the standpoint of user and function. Note that the initial seat position depends on the prior transaction.

2.3. Cost.

Raising or lowering the toilet seat is an inconvenience, and our definitions of cost and fairness seek to codify this. There are a number of definitions we could adopt. We could count (i) the total number of position changes (raisings and lowerings), reflecting the notion that the physical act of moving the seat is a burden, or (ii) whether or not any flips are needed (but not how many), perhaps reflecting a one-time hygenic cost per transaction, or (iii) only whether a raising is necessary, since lowering can be accomplished using a foot and gravity and a loud crash without the need to touch the seat.

We'll adopt (i) and consider each raising or lowering of the seat to be of equal cost. To be precise, the cost of a transaction is the number of seat position changes needed. This will be 0, 1, or 2.

2.4. **Fairness.** In broad terms, we want a strategy which ensures that M and F each incur the same cost. However, there are various possible ways to define this. Given some fixed prior probability distribution (PD) over the initial seat position, we could demand that the expectation value of every individual transaction be identical for M and F. However, this is unnecessarily restrictive. Moreover, the relevant PD evolves based on the strategy.

A better approach is to require (i) that the mean costs for M and F converge in probability to the same value and (ii) that they do so independent of the initial condition (i.e. the seat position prior to the first transaction). Put simply, over time the two participants experience the same average cost. This will, in fact, prove tantamount to demanding equality for individual transactions, but only for a particular PD: the stationary distribution of a certain Markov Chain.

2.5. **State.** In a given transaction, the user has three binary pieces of information: (i) the initial seat position: U or D, (ii) their own anatomy: M or F, and (iii) the function they need to perform: #1 or #2.

These completely define the states, of which there are 8. We'll write a specific state as XYZ, where X is U or D, Y is M or F, and Z is #1 or #2.

Where numeric indices are preferable, we'll use the value 0 for D, M, #1, and the value 1 for U, F, #2. As such, a state string can be written as a binary number (ex. 010) with decimal (or octal) values $0-7$.

2.6. **Strategy.** In a given transaction, a participant enters the restroom, adjusts the seat if necessary, does their business, and leaves the toilet seat U or D. This latter decision is determined by their strategy, a prescription based on the information available to them and perhaps a random element.

We wish to keep things simple for the participants. At the very least, we don't want to require that they track history, tallying past transactions, etc.

Ideally, they always could do the same deterministic thing without consulting the state or employing any random element. For example, they could leave the seat U or D, or maybe leave it as used. Unfortunately, no fair solution exists along these lines. Use of the state is necessary.

The next best scenario would be a simple deterministic strategy that does depend on the state. There are 8 states and thus $2^8 = 256$ deterministic strategies, which we can label $0-255$ based on their binary values. Each consists of 8 binary variables telling the user whether to leave the seat U or D after encountering a given state. The aforementioned always-U or always-D stateless examples would correspond to strategies 0 and 255 (in binary, all 0's or all 1's).

Note that M and F each only employ 4 of the strategy variables. However, we cannot assume symmetry (which would reduce consideration to 16 strategies). It is quite possible that a fair strategy will require different choices by M and F for the same seat position and function.

Though there are 256 deterministic strategies, the presence of $0 < p < 1$ seems to speak against the likelihood any would be fair in general. Even were we to find a deterministic fair strategy, our work would not be done. We want the best fair strategy, one which minimizes the overall average cost subject to our fairness constraint. To this end, we still must consider randomized strategies.

Because each decision is binary, any complicated set of random variables and calculations can be reduced to a simple Bernoulli random draw with some parameter x. However, the choice of x can be a function of the state.

A strategy thus consists of 8 distinct numbers $x_i \in [0, 1]$, one for each possible state. Of course, each participant only makes use of 4. Suppose M needs to do #1 and finds the seat D. He raises the seat (incurring one unit of cost), does his thing, and then flips a weighted coin with Bernoulli parameter x_{DM1}. With probability x_{DM1} he leaves the seat D, incurring another unit of cost. His average incurred cost for the transaction is $1 + x_{DM1}$.

Note that the 8 parameters behave like probabilities but are independent of one another. They needn't sum to 1 and are not constrained in any way. A deterministic strategy corresponds to each x_i being 0 or 1.

2.7. **Flips.** In our analysis, we can treat seat flips in two ways. We could (i) track absolute positions, and then determine flips as changes in position or (ii) count flips directly and then determine changes in position based on those flips. For example, we could speak of a decision to flip the seat or a decision to leave the seat U or D. The two approaches are equivalent, and this simply is a matter of bookkeeping. Certain parts of the calculation are easier one way, and certain parts the other. We'll adopt the absolute approach and speak of a decision to leave the seat U or D. Whether this entails a flip depends on the post-use state of the seat.

2.8. **Toilet Seat.** There are three positions of interest in a given transaction: (i) the initial position on entry, (ii) the necessary position for the function, and (iii) the final position.

Why would the final position (iii) ever differ from (ii)? At first glance, it seems silly to adjust the seat after use. The overall cost only can increase by doing so. Suppose M does #1. The seat now is U. The next user may or may not require initial position D. If fairness is not a concern, then it does not matter who raises or lowers the seat. By lowering it prematurely, there is a chance of incurring an extra two units of cost.

The only reason to consider a final flip, and the raison d'etre of this paper, is to ensure fairness. To do so, we must increase the overall average cost to participants.

2.9. **Optimization.** Ensuring fairness isn't enough, however. As mentioned, it is quite possible to have several fair strategies, some of which have a higher average cost than others.

Fairness always comes at a price. The overall (unfair) cost-minimizing strategy will have a lower average cost than any fair strategy. If the lowest-cost strategy overall has cost 1 for M and 2 for F but the lowest-cost fair strategy has cost 3 for both, it is legitimate to debate the merits of fairness when it drags both participants down. However, such social and philosophical considerations are beyond the scope of this paper. We're just here to calculate.

We must choose how to implement the fairness requirement. We could treat it either as a hard constraint or a soft one. As a hard constraint, we would minimize cost on the fairness-respecting surface. As a soft constraint, we would minimize an unfairness-penalized cost function on the entire space.

For example, suppose C_M and C_F are the average costs for the two participants as functions of the eight x parameters (note that C_M and C_F each depend on all eight x's, even though M and F each only make use of four x's). A hard constraint would minimize $C_M + C_F$ subject to $C_M = C_F$, while a soft constraint would minimize something like $(C_M + C_F) + \lambda |C_M - C_F|^2$, penalizing us for unfairness in some fashion. λ would be pre-specified, and other functional forms are possible depending how aggressively we wish to punish unfairness (at the price of increasing overall cost).

The soft constraint approach is interesting, and in some ways more tractable, but we'll focus on the hard constraint in this paper. The latter requires no a priori choice of functional form or elasticity parameter. It also ensures we'll get a fair solution rather than one which "kind of is fair unless it costs too much." After all, can one place a price on fairness?

2.10. **Summary of Problem.** The problem in its totality can be stated as follows. We're given probability p of function #1, we assume that M and F are equally probable, and we assume that the user and function are drawn independently for each transaction.

Defining the cost of a transaction as the number of changes in position of the toilet seat, and defining the state S to consist of the choice of user, choice of function, and initial seat position, we seek a choice of 8 Bernoulli parameters x_S which minimize the average overall cost subject to the constraint that the average cost for each user is the same.

3. Markov Chain Analysis

It turns out that Markov chains are the right tool for the job. Though the setup smacks of game theory, its cooperative nature makes it less suited to such analysis. It is possible to study the problem as a Markov decision process (MDP) using the Bellman equations, but there is no need. An ordinary Markov Chain (MC) suffices and is more direct. However, we must take care to construct this MC correctly.

Both our strategy and our cost function depend solely on the initial state and the final seat position. The choices of gender and function for each transaction are IID and external, so we cannot naively view the MC in terms of transitions between our 8 states. We can construct a suitable transition matrix, but care must be taken.

It may be tempting to go to the opposite extreme and construct a two-state MC, using U and D as the states and building the gender and function into the transition matrix entries. This is the most intuitive approach, because our cost and fairness functions literally count changes in seat position. We could indeed construct such a 2×2 transition matrix, with each entry a sum over various terms (ex. $U \to U$ would consist of 16 terms). However, this approach is too coarse for our purposes. It does not allow us to compute the separate average costs for M and F, each of which picks certain terms from each of the matrix entries.

The correct approach is to use our 8-state definition but with a twist. As a source, our state makes perfect sense. The transaction occurs in the particular context of that state. By randomly sampling using x_S for the state, the user determines the final seat position. But what does it mean to transition to a final state of our type? The final state of a transaction is the initial state of the next, and the latter involves a gender and function which will be determined at some future point. The trick is to pre-sample those, and treat them as part of the transition. I.e., we line up the next transaction as far as the MC goes. Let's now do this formally.

3.1. **Transition Matrix.** In what follows, I_n denotes the $n \times n$ identity matrix and $diag(\dots)$ denotes the diagonal matrix with specified diagonal elements. We'll consider vectors and matrices with indices running from 0 to 7, corresponding to states $0 = DM1$, $1 = DM2$, $2 = DF1$, $3 = DF2$, $4 = UM1$, $5 = UM2$, $6 = UF1$, $7 = UF2$. We'll assume $0 < p < 1$ (otherwise there's no point to the analysis because there's only one bodily function). x_i denotes the strategy's Bernoulli probability of leaving the final seat position D when the user encounters state i, and we'll use x to denote $(x_0 \dots x_7)$ collectively.

Denoting the transition matrix T, the transition probability T_{ij} between states $i \to j$ consists of three independent components: (i) a Bernoulli draw based on x_i to determine the final seat position (i.e. the next round's initial seat position), (ii) a 50-50 draw to determine the next round's user (M or F), and (iii) a Bernoulli draw based on p to determine the next round's function (#1 vs #2). Notably, the latter two are independent of the state i. Ex. $T_{00} = \frac{p}{2}(1 - x_0)$.

We can write the transition matrix T as $T = T^F T^U T^P$, where T^F is the function-related component, T^U is the user-related (i.e. M or F) component, and T^P is the position related component. $T^U = 0.5 I_8$ is constant, $T^P = diag(p, 1-p, p, 1-p, p, 1-p, p, 1-p)$ depends on p, and T^P, the only part which depends on x, consists of four rows of x and four rows of $1 - x$:

$$T^P = \begin{pmatrix} x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ x_0 & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 1-x_0 & 1-x_1 & 1-x_2 & 1-x_3 & 1-x_4 & 1-x_5 & 1-x_6 & 1-x_7 \\ 1-x_0 & 1-x_1 & 1-x_2 & 1-x_3 & 1-x_4 & 1-x_5 & 1-x_6 & 1-x_7 \\ 1-x_0 & 1-x_1 & 1-x_2 & 1-x_3 & 1-x_4 & 1-x_5 & 1-x_6 & 1-x_7 \\ 1-x_0 & 1-x_1 & 1-x_2 & 1-x_3 & 1-x_4 & 1-x_5 & 1-x_6 & 1-x_7 \end{pmatrix}$$

When building T from components like this, we must be careful. We are not applying 3 successive transition matrices, which would be tantamount to allowing transition after transition after transition. Rather, we are building the effect of T from three parts.

Note that, although T^U and T^F obviously commute, T^F does not commute with T^P. Why apply it after, rather than before, T^P? If we apply it before, we would be modifying the initial state and then performing our strategy based on those modified probabilities. This is not what we want. We first apply the strategy, which modifies the state. Only then do we apply the probabilities of the next batter up. So, T^P must appear on the right in the expression for T.

3.2. **Stationary State.** To meaningfully compute costs, we must average over a long sequence of transactions. Moreover, the result generally will depend on the initial state of the chain. The existence of a unique stationary state addresses both of these issues. In that case, we are guaranteed the stationary state will be arrived at in a finite number of steps, regardless of the initial state. Once in the stationary state, all transactions statistically look the same, and the average costs over the chain converge to the average costs for that single transaction.

We actually need only concern ourselves with stationarity in U/D (i.e., $\sum_{i=0,1,2,3}(Tv)_i = \sum_{i=0,1,2,3} v_i$ and $\sum_{i=4,5,6,7}(Tv)_i = \sum_{i=4,5,6,7} v_i$. But since the other components (T^U and T^F) are independent and constant, the effect is the same. Stationarity of the 2×2 MC is equivalent to stationarity of the 8×8 MC.

3.2.1. *Uniqueness of Stationary State.* In order for a MC to have a unique stationary distribution, it must be irreducible, meaning that every state is reachable from every other in a finite number of steps. T trivially is irreducible in almost all cases, because any state can be reached from any other in a single step unless certain of the x_i's are 0 or 1.

There are two key banks of states: $0-3$ are D and $4-7$ are U. As long as some transitions take U to D and D to U, we can get between the two banks of states. Within each bank of 4 states, T^U and T^F allow transitions. The only problem cases are those in which D states only transition to D states or U states only transition to U states. In those cases, once we land in U or D we cannot get out.

Specifically, our non-irreducible cases are of the form: (i) $x = (x_0, x_1, x_2, x_3, 0, 0, 0, 0)$ and (ii) $x = (1, 1, 1, 1, x_4, x_5, x_6, x_7)$. In those cases, the problem can be reduced to a simpler MC confined to the 4 states of the appropriate bank. However, the resulting strategies are uninteresting. Consider (i). In the reduced four-state (all U) problem (and when $0 < x_0, x_1, x_2, x_3 < 1$), the stationary state clearly is $(p/2, (1-p)/2, p/2, (1-p)/2)$ and all strategies reduce to $x_0 = x_1 = x_2 = x_3 = 0$ once in that state. I.e., there is no room to find a fair strategy.

There are three overlap cases, which fit both (i) and (ii), and which we'll examine below: always-up ($x = (0, 0, 0, 0, 0, 0, 0, 0)$), always-down ($x = (1, 1, 1, 1, 1, 1, 1, 1)$), and left-as-found ($x = (1, 1, 1, 1, 0, 0, 0, 0)$). The always-up strategy reduces to (i), the always-down strategy reduces to (ii), and the left-as-found strategy is a direct sum of the two, reducing to (i) or (ii) based on whether the seat is U or D at the very beginning.

Except where otherwise stated, we will assume we are not in one of the non-irreducible situations. Note that reducibility and rank are distinct concepts, and T has rank ≤ 2.

3.2.2. *Calculation of Stationary Distribution.* We thus have a unique stationary distribution. It can be computed as the (unique) eigenvector of T which has eigenvalue 1. The general eigenvalue equation is $T^F T^U T^P v = \lambda v$, so we solve $T^F T^U T^P v = v$, which may also be written $T^F T^P v = 2v$.

Rather than solve this directly, we can simplify things considerably by using the various constraints. v is a PD, so $\sum v_i = 1$. T^P has only two independent rows, and $T^P v$ produces only two distinct values. T^F turns these into 4 distinct values but only by alternately multiplying by p and $1-p$. The resulting constraints are that $v_2 = v_0$, $v_3 = v_1$, $v_6 = v_4$, $v_7 = v_5$, $v_1 = \frac{1-p}{p} v_0$, and $v_5 = \frac{1-p}{p} v_4$. Also, since $P(\#1) = p$, we need $v_0 + v_2 + v_4 + v_6 = p$.

Designating the elements of v via $2v = (a, b, a, b, c, d, c, d)$, the nontrivial constraints correspond to $b = (1-p)a/p$, $d = (1-p)c/p$. $c = p - a$, and $d = (1-p)(p-a)/p = (1-p) - a(1-p)/p$. Defining $z \equiv (1-p)/p$ and noting that $1 - p - az = (p-a)z$, we have $2v = (a, az, a, az, (p-a), (p-a)z, (p-a), (p-a)z)$.

To solve this, we must apply the stationary state condition $Tv = v$. But we need only do so for a single row since a is the only degree of freedom and the rest follows by construction. Using the first row, we get $(p/2)(v \cdot x) = v_0 = a/2$, which can be expanded and solved for a in terms of x. The solution can be written $a = N_1/D$, where $N_1 \equiv p^2(x_4+x_6)+p(1-p)(x_5+x_7)$, and $D \equiv 2 - p(x_0 + x_2 - x_4 - x_6) - (1-p)(x_1 + x_3 - x_5 - x_7)$. It is not hard to see that $p - a = N_2/D$ where $N_2 \equiv 2p - p^2(x_0 + x_2) - p(1-p)(x_1 + x_3)$. From this, we have $v = (N_1, zN_1, N_1, zN_1, N_2, zN_2, N_2, zN_2)/(2D)$.

It is easy to check that (i) $\sum v_i = 1$, (ii) $P(M) = P(F) = 0.5$, and (iii) $P(\#1) = p$, as they must by construction. We also can compute the derived quantity $P(D) = [a+az+a+az]/2 = a(1+z) = a/p$.

3.3. **Cost and Unfairness.** By using an 8×8 transition matrix, not only can we compute the total average cost C, but also C_M and C_F, the average costs seen by M and F, as well as $u = (C_M - C_F)/2$, the unfairness. Clearly, $C = (C_M + C_F)/2$. The exact scale doesn't matter for our purposes, and we could just as well work with $2C$ and/or $2u$.

To compute C_M, C_F, C, and u, we begin with a cost vector c. c_i is the mean cost seen by a user who encounters state i. The number of flips depends only on the initial state and the final seat position, so every c_i has two terms (which may be 0). For our definition of cost, $c = (1 + x_0, 1 - x_1, 1 - x_2, 1 - x_3, x_4, 2 - x_5, 2 - x_6, 2 - x_7)$.

By definition, $C_M = 2\sum_{i=0,1,4,5} v_i c_i$ and $C_F = 2\sum_{i=2,3,6,7} v_i c_i$. Note that these are average costs to M and F rather than the contributions to C, so we must divide by $P(M) = P(F) = 0.5$. Hence the factor of 2.

Let's consider some simple strategies. Note that the first three are non-irreducible cases we touched on earlier, and not amenable to our stationary state approach.

3.4. **Simple case: always down.** One simple deterministic strategy is to always leave the seat down. In this case, T is not irreducible (the U states cannot be reached from any other), but it's easy to see that the corresponding 4×4 calculation yields a stationary state of $v = (p/2, (1-p)/2, p/2, (1-p)/2)$, with $C_M = 2p$ and $C_F = 0$, so $C = p$ and $|u| = p$.

3.5. **Simple case: always up.** Similarly, we could always leave the seat up. Again, T is not irreducible, and we solve a corresponding 4×4 calculation to get a stationary state of $v = (p/2, (1-p)/2, p/2, (1-p)/2)$ (now for the U states), with $C_M = 2(1-p)$ and $C_F = 2$, so $C = 2 - p$ and $|u| = p$.

3.6. **Simple case: left as found.** Another simple deterministic strategy is to leave the seat in the position in which it was found, corresponding to $x_0 = x_1 = x_2 = x_3 = 1$ and $x_4 = x_5 = x_6 = x_7 = 0$. Once again, T is not irreducible. However, in this case the result is entirely dependent on the very first position of the toilet seat (since every subsequent initial position will be the same). We'll refer to these as U-chains and D-chains, and they correspond directly to the "always up" and "always down" strategies just discussed.

3.7. **Simple case: unconstrained minimum cost.** When fairness is not a concern, the minimal cost is achieved by the deterministic strategy of performing no final flip (i.e. leaving the seat in the position dictated by the function). This corresponds to $x_0 = x_4 = 0$, and $x_1 = x_2 = x_3 = x_5 = x_6 = x_7 = 1$. The seat is left down unless M did #1. As discussed earlier, it is obvious this is optimal because no flip is made until necessary.

In this case, $D = 2$, $N_1 = p^2+2p(1-p) = 2p-p^2 = p(2-p)$, and $N_2 = 2p-p^2-2p(1-p) = p^2$. From this, $v = (p(2-p), (1-p)(2-p), p(2-p), (1-p)(2-p), p^2, p(1-p), p^2, p(1-p))/4$. The costs are $c = (1, 0, 0, 0, 0, 1, 1, 1)$, from which we have $C_M = (p(2-p) + p(1-p))/2 = (3p - 2p^2)/2$ and $C_F = (p^2 + p(1-p))/2 = p/2$. The overall cost is $C = p - p^2/2$ and the unfairness is $p/2 - p^2/2 = p(1-p)/2$.

3.8. **A simple, fair strategy.** There are no deterministic fair strategies in general. Let us consider the simplest non-deterministic strategies — ones in which all the x_i are equal. We'll denote by x' the common value of the x's. Then $D = 2$, $N_1 = 2px'$, and $N_2 = 2p(1-x')$. From this, $v = (px', (1-p)x', px', (1-p)x', p(1-x'), (1-p)(1-x'), (1-p)x', (1-p)(1-x'))/2$.

It is easy to calculate $C_M = 2(1-p) + 2x'(2p-1)$ and $C_F = 2(1-x')$. Fairness requires that $1 - p + x'(2p-1) = (1-x')$, which reduces to $p = 2px'$, whose solution is $x' = 1/2$ independent of p.

Thus the simplest possible random strategy is fair. Every user flips an unweighted coin and leaves the seat U or D according to the outcome. Even without our analysis, it is obvious this is fair. The previous seat position is random, and there is a 50-50 chance any new user will need to flip it.

Though this simple strategy is fair, it is far from optimal. $C = C_M = C_F = 1$, and both users are worse off than with the optimal unfair strategy. We achieve fairness, but at significant cost to both M and F. Can we do better?

Let us now construct an optimal fair strategy. It too will incur a cost over the optimal unfair strategy, but a considerably smaller one.

4. Optimal Fair Strategy

To perform our constrained minimization, we must compute C_M and C_F, from which $C = (C_M + C_F)/2$ and $u = (C_M - C_F)/2$.

To minimize C subject to $u = 0$, we use a Lagrange multiplier and solve the set of equations $\frac{\partial(C-\lambda u)}{\partial x_i} = 0$ along with $u = 0$ (verifying that the $2nd$ derivatives are positive definite at any solution point, of course). The solutions we find must lie in $0 \leq x_i \leq 1$ and avoid the aforementioned non-irreducible cases. As always, we assume $0 < p < 1$.

It turns out there is no solution to this set of 9 equations in the relevant domain. That does not mean we are out of luck, just that no local minimum exists in the $u = 0$ subspace. Let us proceed as usual and see what breaks down. Along the way, we will encounter two situations which we'll refer to as "irrelevant problem cases": $x_1 = 1/(1-p)$ and $x_5 = p/(p-1)$. They are "irrelevant" because neither can occur for $x_1, x_5 \in [0, 1]$ and $p \in (0, 1)$.

Define $f_i(x_0, \ldots, x_7, \lambda) \equiv \partial_i C - \lambda \partial_i u$ for $i = 0 \ldots 7$ and $f_8(x_0, \ldots, x_7, \lambda) \equiv u$. Our 9 equations now can be written $f_i = 0$. We begin by solving $f_8 = 0$ (i.e. $u = 0$) for x_7 in terms of the other x's and λ. Except for the irrelevant problem case $x_1 = 1/(1-p)$, there is no obstruction to doing so. Plugging the result into the remaining 8 equations, the f's now are functions of $(x_0, \ldots, x_6, \lambda)$, and $f_8 = 0$. From the ratios of the equations, it is apparent that several are linearly dependent except in the irrelevant problem cases. Specifically, $f_0 \propto f_4$, $f_1 \propto f_5$, and $f_2 \propto f_3 \propto f_6 \propto f_7$. I.e., only 3 of the equations are linearly independent. This tells us that the derivative in several directions (not necessarily along specific x-axes) will be 0.

Next, we solve for λ using equation $f_7 = 0$. This can be accomplished unless $x_5 = x_1 + (p+1)/(p-1)$, which cannot happen in our domain (it would require $x_5 \leq 0$ because $(p+1)/(p-1) < -1$ for $p \in (0,1)$). Plugging the result into the remaining 7 equations, the f's now are functions of $(x_0, \cdots, x_6)$, and both $f_7 = 0$ and $f_8 = 0$. However, $f_2 = f_3 = f_6 = 0$ as well because those were proportional to f_7. Of the remaining 4 f's, only two are linearly independent. We can take these to be f_0 and f_1.

There is no solution to $f_0 = 0$ and $f_1 = 0$, but all is not lost. f_0 depends only on x_1 and x_5, and $f_0 > 0$ for all $x_1, x_5 \in [0,1]$. I.e., the derivative $\frac{\partial(C-\lambda u)}{\partial x_0} > 0$. We've already constrained ourselves to $u = 0$, so C is a monotonically increasing function of x_0 on the relevant domain. I.e., the minimum cost on the 7-dimensional $u = 0$ surface must lie on the $x_0 = 0$ edge. The same holds for $f_4 = 0$, so we set $x_4 = 0$ as well.

Plugging in $x_0 = x_4 = 0$, the remaining functions f_1 and f_5 (which also depend only on x_1 and x_5) both are negative on the domain defined by $x_0 = x_4 = 0$ and $u = 0$. I.e., C is minimized on the $x_1 = x_5 = 1$ edge.

We now have ascertained that, on the $u = 0$ surface, C is minimized when $x_0 = x_4 = 0$ and $x_1 = x_5 = 1$. We'll denote this subspace of the $u = 0$ surface Y. All the other derivatives are 0 on Y, f_7 and f_8 because we solved for variables using them and f_2, f_3, and f_6 because they are proportional to f_7. On Y, C is independent of x_2, x_3, and x_6 (we already eliminated x_7 to confine us to the $u = 0$ surface).

In other words, the cost is the same for all x_2, x_3, and x_6 once we are in Y. It does not matter what choice of those variables we make, so we may as well keep things simple and pick $x_2 = x_3 = x_6 = x_7$. I.e., we declare that all four of those variables are equal. When we solved $u = 0$, we obtained an expression of the form $x_7 = g(x_0, \ldots, x_6)$. We now must solve $x_7 = g(0, 1, x_7, x_7, 0, 1, x_7)$, which yields $x_7 = (1+p^2)/(1+p)$.

Let's step back and consider what has happened. Confining ourselves to the $u = 0$ surface, there is no local minimum. In fact, C is a monotonic function of x_0, x_1, x_4, and x_5. We thus must confine ourselves to the $x_0 = x_4 = 0$, $x_1 = x_5 = 1$ edge of our surface, which we called Y. On Y, C is constant, so we can pick x_2, x_3, and x_6 as we please. Each choice of these defines a unique x_7 in Y. I.e., if we think of x_2, x_3, x_6, and x_7 (each with $0 \leq x_i \leq 1$) as a $4d$ box, Y is a surface above the x_2, x_3, x_6 axes. This is easier to visualize in $3d$ (if we ignored x_2, then x_7 would be the z axis and x_3, x_6 the $x-y$ axes and Y would be an ordinary surface above the $x-y$ plane). $x_2 = x_3 = x_6 = x_7$ defines a line in the $4d$ cube, running from the origin to $(1,1,1,1)$. This must intersect the surface Y somewhere in that unit cube, so we are guaranteed a solution to $x_2 = x_3 = x_6 = x_7$. It is conceivable it could intersect more than once, but that is not the case here.

The net result is there are many cost-minimizing fair strategies. All involve deterministic behavior by M: that he leave the seat as he used it (U for #1 and D for #2). If F did the same, we would have the unconstrained optimal solution, so she has to do a little work (and sacrifice a little) in order to be fair to M. However, we can simplify life for F by making her strategy uniform. This is possible because we have so much freedom to choose the remaining x's. All F must do is flip the same weighted coin each time, leaving the seat D with probability $(1+p^2)/(1+p)$.

This strategy has $u = 0$ by construction and $C = C_M = C_F = p(2-p)/(1+p)$, with a maximum cost of around 0.536 at $p = \sqrt{3} - 1$. Note that our earlier fair strategy ($x_i = 1/2$

for all i) is not in Y. It lies on the $u = 0$ surface (as all fair strategies must) but not along the $x_0 = x_4 = 0$, $x_1 = x_5 = 1$ edge.

How much does fairness cost us relative to the unconstrained minimum? The difference for C_M is $\frac{2p^3-3p^2+p}{2(p+1)}$, which is positive for $p < 0.5$ and negative for $p > 0.5$. I.e., if #1 is more likely than #2 then M benefits from the fairness strategy in absolute terms (not just relative to F). The difference for C_F is $\frac{3(p-p^2)}{2(p+1)}$ which always is positive and has a maximum of around 0.26 at $p = \sqrt{2} - 1$. F always must sacrifice to make things fair, but if $p < 0.5$ then M and F both come out behind. In that case, M may decide it is best to forego fairness, allowing F to reduce her cost more than he reduces his.

Note that the optimal fair strategy is a significant improvement over our simple random fair strategy with all the $x_i = 0.5$. The improvement varies from a cost savings just under 1 to around 0.464, and is felt by both participants equally (since both strategies are fair).

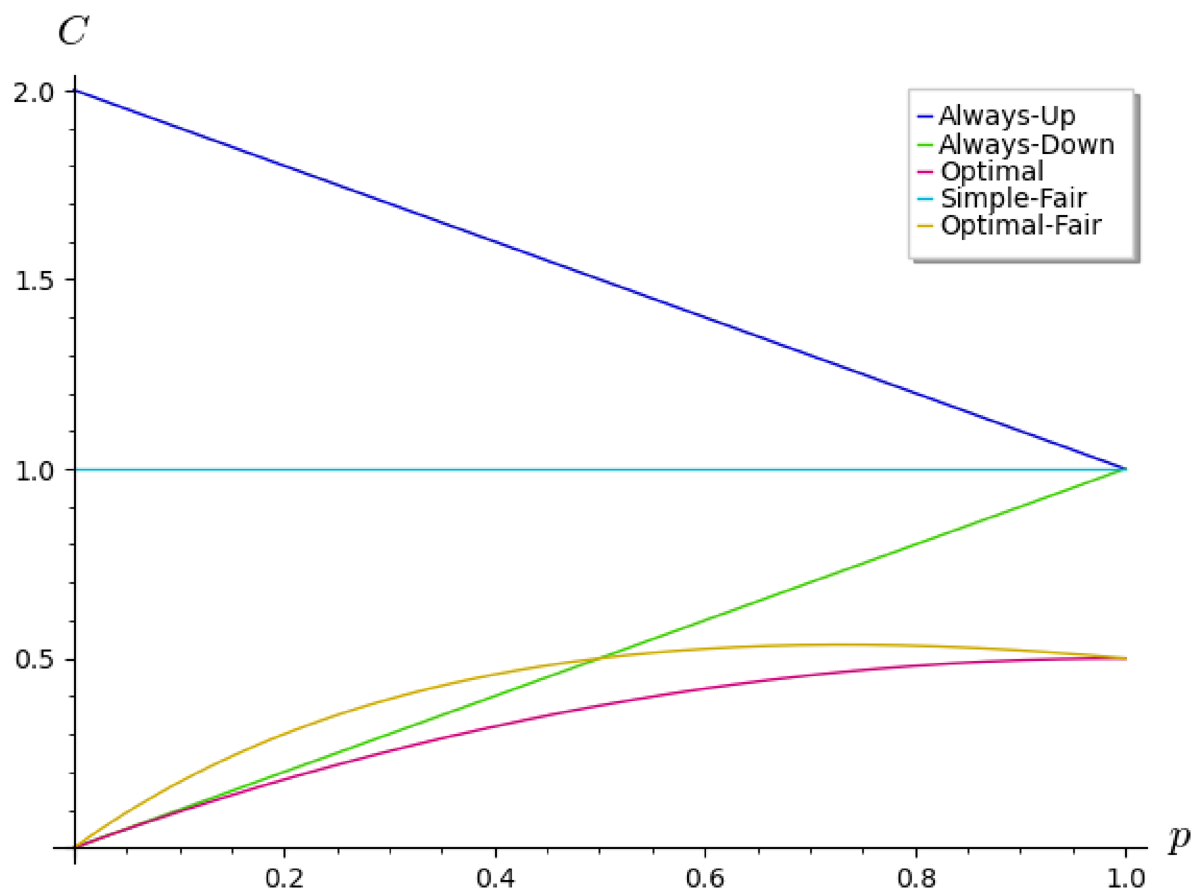
5. Summary of Results

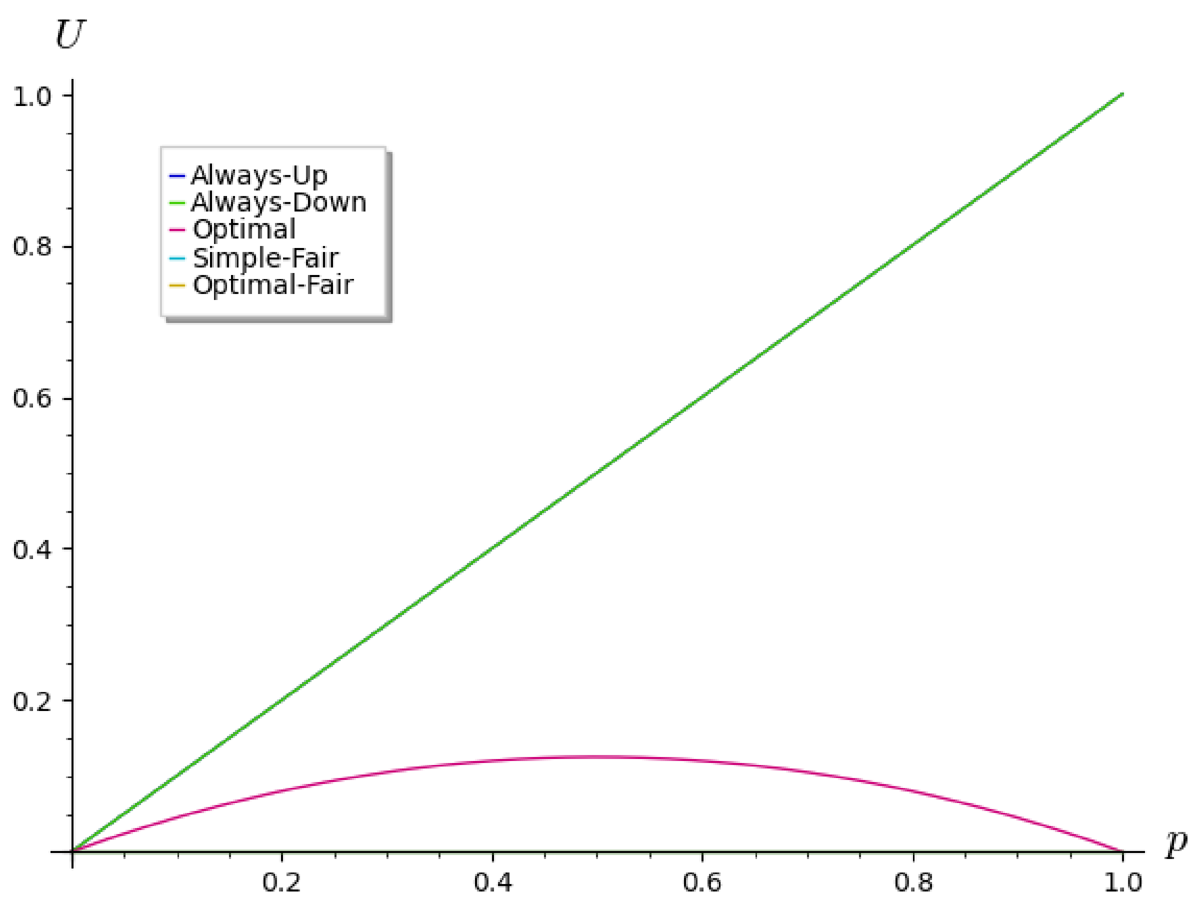
First, let's summarize the results from the various strategies we discussed. The strategies are as follows:

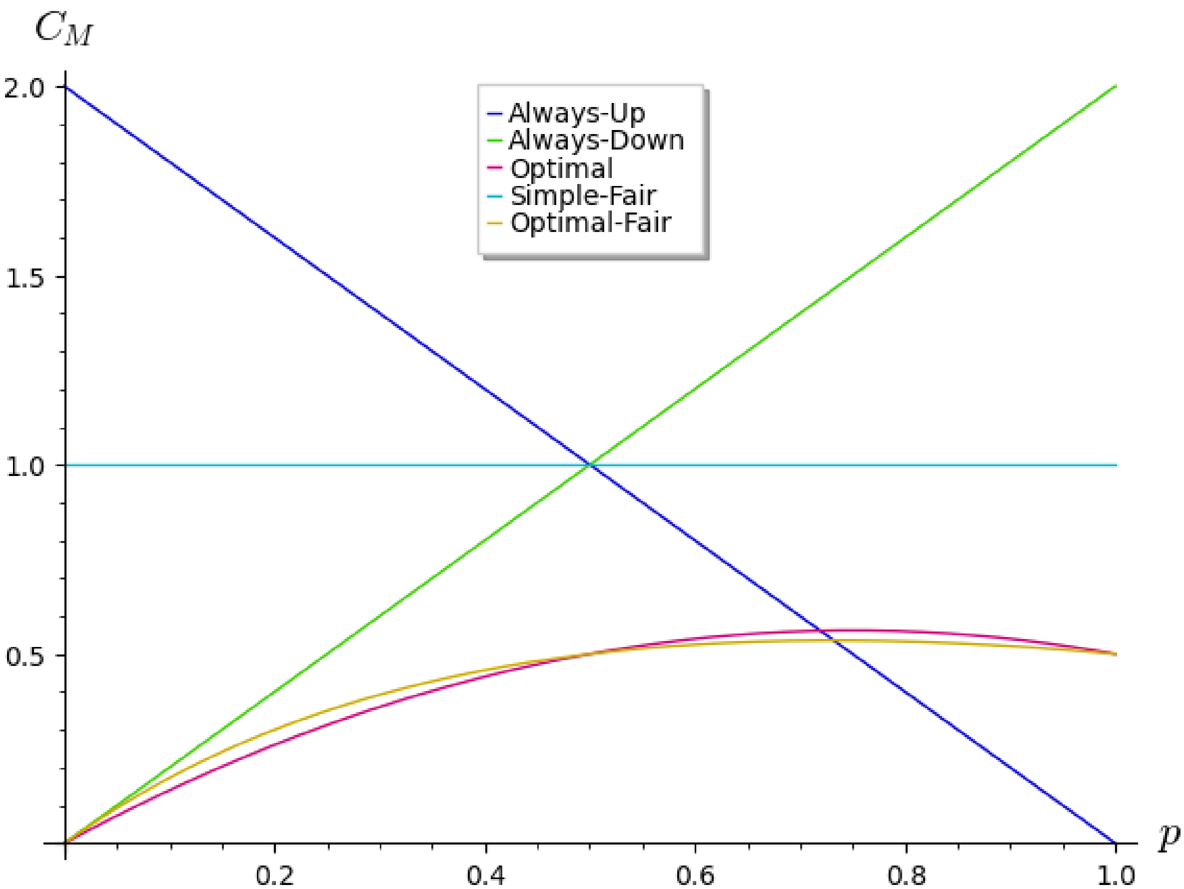
- Always-Up: Both M and F leave the seat up when done.
- Always-Down: Both M and F leave the seat down when done.
- Optimal: The optimal strategy when fairness is not an issue. Both M and F leave the seat in the position required for use. No final flip is involved.
- Simple-Fair: A non-optimal fair strategy in which both M and F always flip a coin when done and leave the seat up or down with 50-50 probability.
- Optimal-Fair: Our optimal fair strategy (one of many), in which M always leaves the seat in the position required for use and F leaves it down with probability $\frac{1+p^2}{1+p}$.

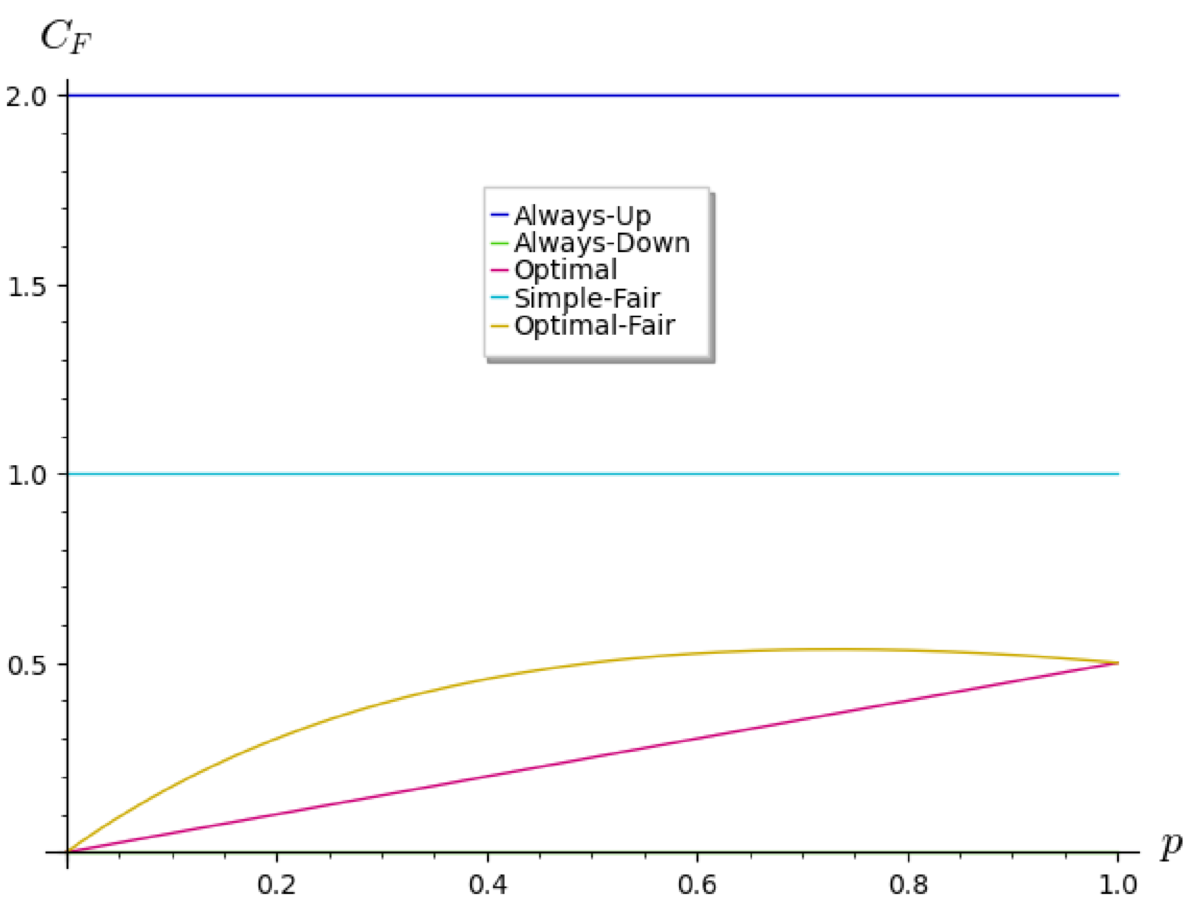
In the following table, $P(D)$ denotes the probability that a user finds the seat down in the stationary state.

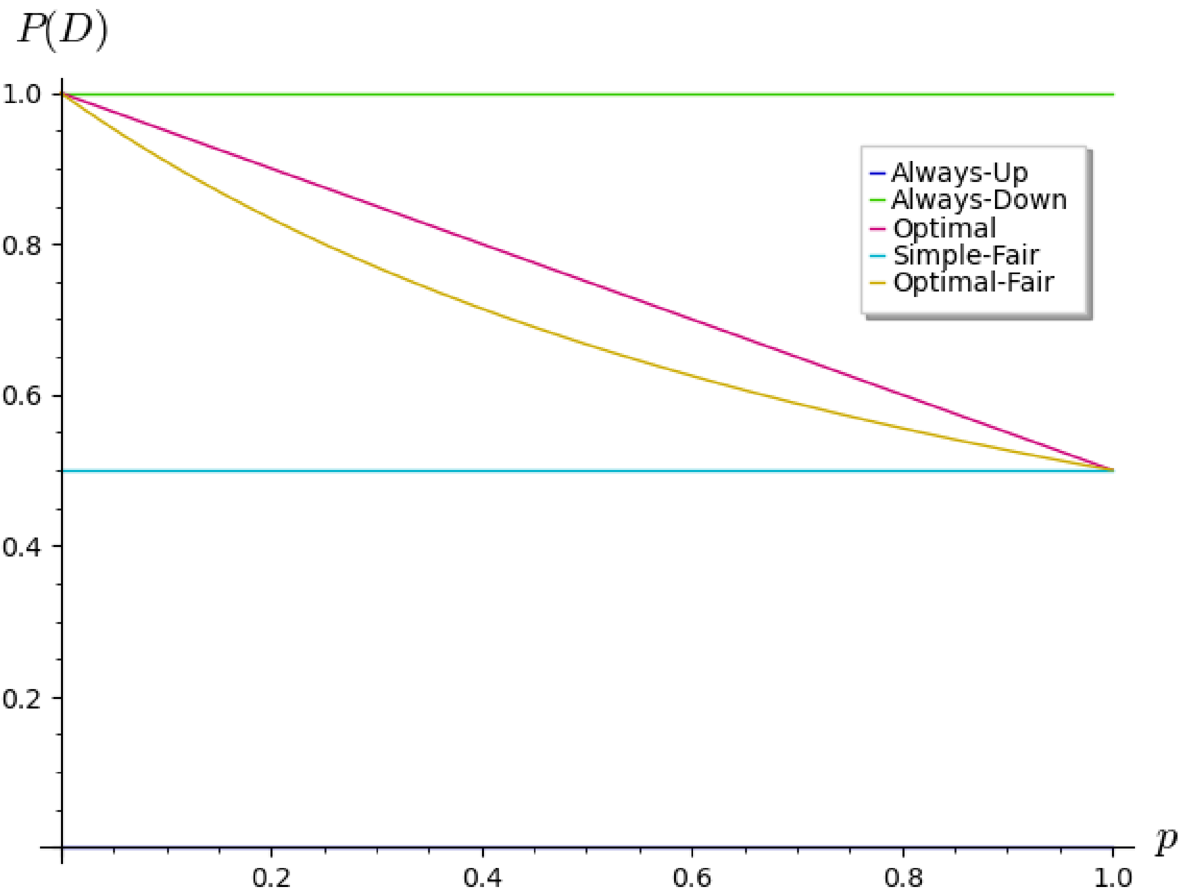
Strategy	C	$\lvert u\rvert$	C_M	C_F	$P(D)$	x
Always-Up	$2-p$	p	$2(1-p)$	2	0	$(0,0,0,0,0,0,0,0)$
Always-Down	p	p	$2p$	0	1	$(1,1,1,1,1,1,1,1)$
Optimal	$p-\frac{p^2}{2}$	$\frac{p(1-p)}{2}$	$\frac{3p-2p^2}{2}$	$\frac{p}{2}$	$1-\frac{p}{2}$	$(0,1,1,1,0,1,1,1)$
Simple-Fair	1	0	1	1	$\frac{1}{2}$	$(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$
Optimal-Fair	$\frac{p(2-p)}{1+p}$	0	$\frac{p(2-p)}{1+p}$	$\frac{p(2-p)}{1+p}$	$\frac{1}{p+1}$	$(0,1,\frac{1+p^2}{1+p},\frac{1+p^2}{1+p},0,1,\frac{1+p^2}{1+p},\frac{1+p^2}{1+p})$


C
2.0
1.5
1.0
0.5
0.2
0.4
0.6
0.8
1.0
p
Always-Up
Always-Down
Optimal
Simple-Fair
Optimal-Fair


U
1.0
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1.0
p
Always-Up
Always-Down
Optimal
Simple-Fair
Optimal-Fair


C_M
2.0
1.5
1.0
0.5
0.2
0.4
0.6
0.8
1.0
p
Always-Up
Always-Down
Optimal
Simple-Fair
Optimal-Fair


C_F
2.0
1.5
1.0
0.5
0.2
0.4
0.6
0.8
1.0
p
Always-Up
Always-Down
Optimal
Simple-Fair
Optimal-Fair



6. Conclusion

We have our answer. An optimal fair strategy is for M to leave the seat as he used it and for F to leave the seat down with probability $(1+p^2)/(1+p)$. Now that humanity's greatest challenge has been overcome, the species doesn't really have much reason to go on, much less continue to cover the planet in #1 and #2.

Cambridge, MA
Email address: `khalpern@alum.mit.edu`