# On the Advantage of Being the First Server

Refael Hassin
*Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel*

*Dedicated to the memory of my father Simon Hassin, 1909–1995*

The following example illustrates the problem treated in this paper: Two gas stations are located one after the other on a main road. A driver who needs to fill his tank sees the queue situation at the first station but not at the second one. The driver estimates the expected waiting time at the first station, compares it to the *conditional* expected waiting time at the second one, and decides which station to enter. The second station is assumed to be on the driver's route so that no extra cost is involved in choosing it. Is it true that the first station always gets a higher share of the demand than the second one? We model the situation in terms of queueing theory and answer the question.
(*Queues; Threshold Strategies*)

## 1. Introduction

In a common situation, a producer (or a seller) does not provide specific data on the good manufactured (or sold) by him. Instead, the buyer bases his decision to purchase the good on the available statistical properties of the good. This way, there is a common price to all the units with no regard to their true value.[1] This phenomenon may occur even when it is not costly to obtain and store the information for each unit, and to charge different prices for different units. On one hand, there is some loss entailed in selling the better products for the common price. On the other hand, there are gains associated with selling the worse products for this price. The outcome depends on the characteristics of the demand for the good.

In an analogous case, a server may choose not to publicize the number of customers waiting in the queue for service. In such a case, customers base their decision on whether to join the queue on its statistical properties (see, for example, Edelson and Hildebrand 1985). It has been shown in Hassin (1986) that such a strategy, while adopted by a monopolist, may increase both the server's income from selling service and total social welfare.

This note considers the question of whether a non-monopolist server can profit by concealing the queue length. In particular, consider a service system that consists of two servers with separate queues denoted Q1 and Q2. The servers are identical except that the length of Q1 can be observed by an arriving customer, while that of Q2 cannot be observed.

The input to the system consists of a stream of customers. Each wishes to minimize his expected waiting time in the system. Upon arrival, a customer observes the length of Q1, and decides which of the two queues to join.[2] The decision is irrevocable.

As a real-life illustration to our model, consider two gas stations located one after the other on a main road.

[1] Some examples are hotel reservations where the exact room is not specified, theater tickets with no prior seat assignment, and products sold through catalogues and advanced orders. In Glazer and Hassin 1982 it is argued that concealing the exact content of a product sold through subscriptions may increase both profits and social welfare. Racing politicians can also be viewed as sellers of a good, and in Glazer 1990 it is argued that both candidates may want to hide information.

[2] Interesting phenomena are demonstrated in Whitt (1986) and Hlynka et al. (1994). Assuming that both queues can be observed, then in some cases the policy of joining the shortest queue is not a Nash equilibrium in the following sense: given that the other customers join the shortest queue, a given customer may decrease his expected time in the system by sometimes joining the longer queue or by waiting and observing before making his decision.

0025-1909/96/4204/0618$01.25

A driver who needs to fill his tank must choose a station. He sees the queue at the first station while making his decision, but not at the second one. Once he decides to proceed to the second station it is too costly to drive back to the first one. Thus the driver estimates his expected waiting time at the first station and compares it to the *conditional* expected waiting time at the second one. The second station is assumed to be on the driver's route so that no extra cost is involved in choosing it. The next station on that road is sufficiently far away so that a driver who selects the second station will enter it even if he observes, upon his arrival to this station, a long queue there.

Thus, an arriving customer bases his choice on the actual state at Q1, and on the known statistical properties of Q2. These properties are used by the customer to compute the expected length of Q2 conditioned on the length of Q1. Clearly, this conditional expectation takes into account the choice strategies of the other individuals in the population. Therefore, the strategy adopted by any given individual is a function of the strategies used by the others. Below, we will compute a set of strategies that satisfy some stability requirements.

We will investigate threshold strategies of customers, that is, strategies in which customers enter Q1 only if its length is at most some given constant. We will follow Hassin and Haviv (1994) and allow a generalized form of threshold strategies. We will claim that the individual's optimal threshold is monotone nonincreasing with the threshold adopted by the others in the population—the higher is the other customer's tendency to enter Q1, the lower is the tendency of a given customer to do so. This behavior, called *avoid the crowd* in Hassin and Haviv (1994), results in a unique equilibrium threshold strategy.

For a given integer value of the threshold, the system consists of a queue, Q1, with a finite waiting room whose overflow constitutes the input to a secondary queue, Q2. Overflow streams in queues and systems of queues have been analyzed by many authors. For example, Cinlar and Disney (1967), van Doorn (1984), R. Guérin, and L. Y. C. Lien (1990). The results are relatively complex and not directly applicable for the present study in which we allow a more general form of overflow determined by noninteger thresholds. We found it convenient to compute the solutions directly from the state transition equations.

Our goal is to investigate the equilibrium threshold strategies and their effect on the utilization of the two servers. We numerically solve the model with Poisson arrivals and exponential service. Each instance of this model is characterized by a single parameter, the utilization factor of the system, defined as the ratio of the average arrival rate to the system to the average joint service rate of the two servers. We solve for two functions of this parameter, the threshold function and the ratio of the number of customers who join Q1 to those who join Q2.

We find that the equilibrium arrival rate to Q1 is larger than that of Q2. Consequently, we will claim that under the stated model, if the servers have the option of publicizing the queue length at their facilities, both servers will choose to do so.

## 2. The Model

We consider two servers with separate queues, Q1 and Q2. The input process consists of a common stream of customers with an average arrival rate $\lambda$. The service process at each server is exponential with a rate of $\mu$. The utilization factor of the system is $\rho = \lambda/2\mu$. A customer observes the length of Q1, but cannot observe the length of Q2. Upon arrival the customer irrevocably decides which queue to join. His objective is to minimize the expected queueing time.

A customer's *strategy* is a function $s(L) : \mathbb{N} \rightarrow [0, 1]$. Its interpretation is that when the observed length of Q1 is $L$, the customer joins Q1 with probability $s(L)$ and Q2 with the complementary probability, $1 - s(L)$.

A customer takes into account the strategies adopted by others in the population. Therefore, while each customer chooses a strategy that, for any given $L$, minimizes his expected wait, the set of strategies in the population must satisfy some stability conditions. In this note, we consider (symmetric) *Nash equilibrium strategies*. A strategy defines a Nash equilibrium if when every individual in the population follows it, no one can reduce his expected waiting time by deviating from it.[3]

[3] The equilibrium strategies in our case have stronger stability properties. They are *Evolutionary Stable Strategies* in the sense defined in Maynard Smith (1982). This fact follows from more general results presented in Hassin and Haviv (1994).

Customer equilibrium in queueing models have been the subject of several papers. Examples are Edelson and Hildebrand (1985), Hassin (1986), Hassin and Haviv (1994), Haviv (1991), Luski (1976).

In the context of our model, it is natural to consider strategies in which customers enter Q1 if and only if its length is less than a given constant called a *threshold*. However, it is easy to construct instances where, for example, if everyone in the population has a threshold 4 then a deviant whose threshold is 5 has a smaller expected wait, while if everyone in the population adopts a threshold of 5 then deviating to a threshold of 4 reduces the expected wait. (This is the case with the upper function in Figure 2, as will be explained later.) In such a case we may conclude that no strategy of the threshold type defines an equilibrium. Consequently, we follow Hassin and Haviv (1994) and extend the definition of a threshold as follows:

A *threshold strategy* with *threshold* $x = n + r, n \in \mathbb{N}, r \in [0, 1)$, has

$$s(L) = \begin{cases} 1, & L = 0, \ldots, n-1, \\ r, & L = n, \\ 0, & L = n+1, n+2, \ldots. \end{cases}$$

In other words, under a strategy with threshold $x$, a customer always joins Q1 if its size is at most $n - 1$, always joins Q2 if Q1 has more than $n$ customers, and randomizes with probability $r$ when Q1 has exactly $n$ customers.

If $x$ is an integer ($r = 0$), the strategy is *pure*. Otherwise, it is *mixed*. Note that under a pure strategy with threshold $n$, a customer enters the *second* queue whenever he observes $n$ customers in Q1. Thus, in this case, $n$ is the maximum length of Q1. In general, if all the customers in the population follow a threshold strategy with threshold $x$ then the maximum possible number of customers in Q1 is $\lceil x \rceil$.

## 3. The Existence of a Threshold Equilibrium

Consider an individual who assumes that others follow the threshold strategy defined by $x$. For any given length $L$, of Q1, the individual enters Q1 if and only if $L$ is not larger than the *conditional* expected length of Q2.

Let $f(L, x)$ denote the expected length of Q2 given that the length of Q1 is $L$ and that all the individuals in the population follow a threshold strategy defined by $x$. Let $g(L, x) = f(L, x) - L$ denote the expected difference between the lengths of Q2 and Q1 given the length $L$ of Q1 and the threshold $x$. These two functions are defined for all $x \geq 0$ and $L \in \{0, \ldots, \lceil x \rceil\}$. We have investigated $g(L, x)$ numerically. Figure 1 gives a typical case with $\rho = 0.9$ and various $x$ values. Of course, the graphs shown are interpolations of the function at the integer $L$ values. We observe two important properties:

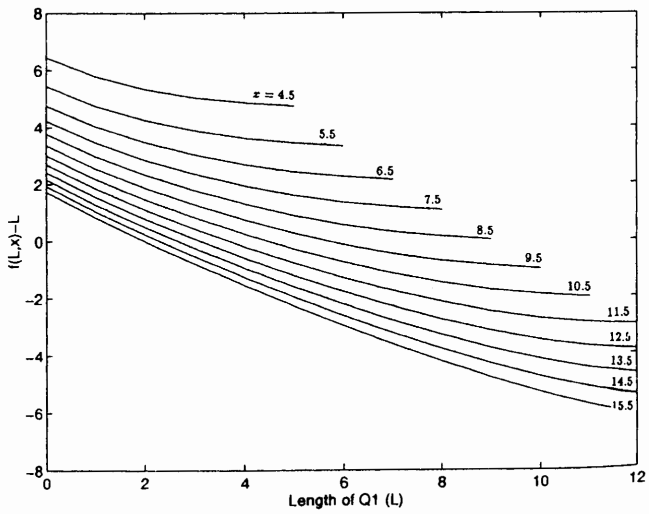
(1) The function is monotone decreasing in $L$;

(2) The function is monotone decreasing in $x$.

The individual customer's optimal response is to enter Q1 if $g(x, L) > 0$, and to enter Q2 if $g(L, x) < 0$. The customer is indifferent between the options if $g(L, x) = 0$. From the monotonicity of the function with respect to $L$ it follows that for a given value $x$, there will be a maximal size of $L$ such that entering Q1 is an optimal response. More exactly, the customer's optimal strategy is of the threshold type: For $k(x) = \min\{L: g(L, x) \leq 0\}$, if $L \in \{0, \ldots, k(x) - 1\}$ join Q1, otherwise join Q2.

From the monotonicity of $g(L, x)$ with respect to $x$, it follows that $k(x)$ is a nonincreasing step function. The discontinuity points of $k(x)$ correspond to thresholds $x$ such that the individual is indifferent between two consecutive thresholds.

**Figure 1** **Expected Difference Q2 − Q1 Conditioned on Q1 (rho = 0.9)**



A Nash equilibrium threshold is a value $x$ such that when adopted by everyone in the population it becomes an optimal choice for each individual. In a mixed equilibrium this means that the individual is indifferent between $\lfloor x \rfloor$ and $\lceil x \rceil$, and therefore he is also ready to randomize between the two values.

In other words, $x$ defines an equilibrium if either $k(x) = x$, or if $x$ is between $k(x-)$ and $k(x+)$ (it is convenient to view both cases as solutions to the equation $k(x) = x$, that is, as fixed points of $k(x)$). In both cases, if all customers adopt strategy $x$, then this is also an optimal strategy for each of them and none has an incentive to deviate from this strategy. In the case of a mixed strategy $x$, then given that the others adopt this $x$, the individual is indifferent between the two thresholds $k(x-) = \lceil x \rceil$ and $k(x+) = \lfloor x \rfloor$.

Figure 2 shows two functions $k(x)$. The lower one intersects the function $f(x) = x$ at $x_1$. Since the steps of $k(x)$ have an integer value it follows that $x_1$ is integer and the equilibrium defined by it is pure. The upper function has $x_2$ as a mixed equilibrium. Given that all follow $x_2$, the individual is indifferent between the thresholds $\lfloor x_2 \rfloor$ and $\lceil x_2 \rceil$. In particular, the individual cannot reduce his expected wait under any observed state by using a strategy different from $x$.

Since an equilibrium lies at the "intersection" of an increasing and a nonincreasing function, it is unique. Specific conditions for a value $x$ to define an equilibrium are stated in the next section.

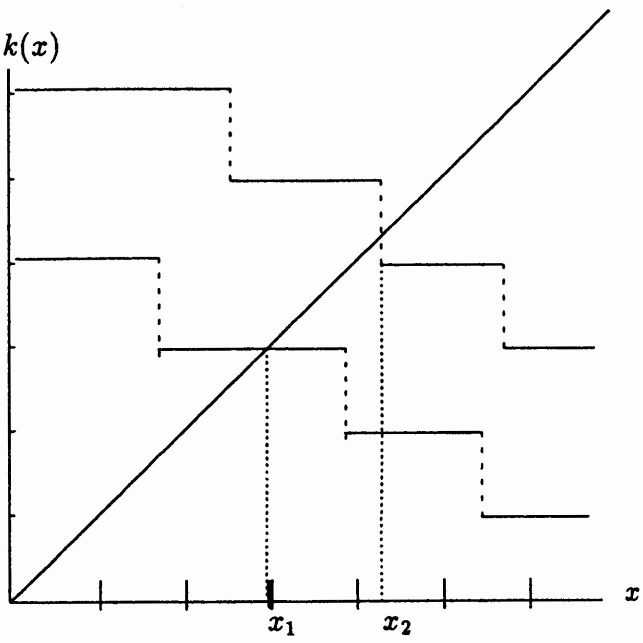
## 4. Conditions for Equilibrium

We denote by $p_{ij}$ the steady state probability of $i$ customers at Q1 and $j$ customers at Q2. These probabilities depend on $x$. However, we simplify the notation by avoiding explicit reference to $x$.

### 4.1. Equilibrium in Pure Strategies

Assume that the entire population uses the integer threshold $n$. In order for $n$ to describe an optimal strategy for an individual, given that all follow this strategy, two conditions are necessary: (1) If a customer arrives and sees $n - 1$ customers in Q1 his optimal choice is to join this queue, and (2) if he sees $n$ customers, his optimal choice is Q2. Moreover, these conditions are also sufficient: If it is optimal to join Q1 when $n - 1$ customers are observed then it is also optimal to do so when less customers are observed. Similarly, if it is optimal to join Q2 when $n$ are observed, then this is also the optimal choice when more than $n$ are observed.

**Figure 2** **Equilibrium Strategies**




Let $q_{ij} = p_{ij} / \sum_{k=0}^{\infty} p_{ik}$ be the probability that the length of Q2 is exactly $j$ given that the length of Q1 is $i$. The equilibrium conditions are:

$$(n - 1) \le f(n - 1, n) = \sum_{j=0}^{\infty} j q_{n-1,j},$$

$$n \ge f(n, n) = \sum_{j=0}^{\infty} j q_{n,j}.$$

Combining the two conditions gives:

PROPOSITION 4.1. *The pure threshold $n \ge 1$ is a Nash equilibrium if and only if*

$$\sum_{j=0}^{\infty} j q_{n,j} \le n \le 1 + \sum_{j=0}^{\infty} j q_{n-1,j}.$$

*In particular, the threshold $n = 1$ is a Nash equilibrium if and only if $\sum_{j=0}^{\infty} j q_{1,j} \le 1$.*

### 4.2. Equilibrium in Mixed Strategies

The condition for the threshold $x = n + r$, $0 < r < 1$, to define a Nash equilibrium is that, given that all follow

**Figure 3** **Threshold and Usage Ratio by Utilization**

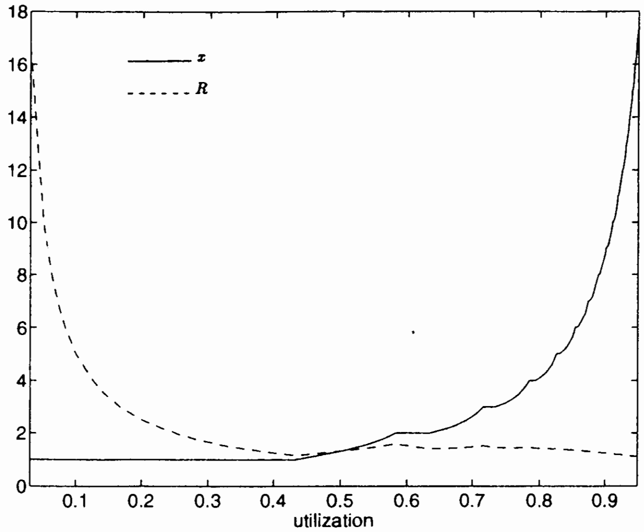
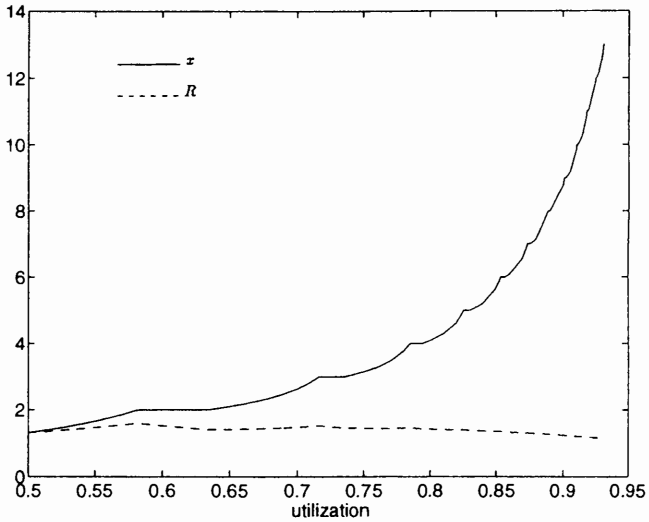



**Figure 4** **Threshold and Usage Ratio by Utilization** (rho > 0.5)




it, an arrival who sees $n$ customers in Q1 will be indifferent between the two options of joining Q1 or Q2.

PROPOSITION 4.2. *The threshold* $x = n + r$, $0 < r < 1$, *specifies a Nash equilibrium if and only if*

$$n = \sum_{j=0}^{\infty} j q_{nj}.$$

## 5. Numerical Solution

Let $\mathbf{I}$ denote the indicator function so that $\mathbf{I}(R) = 1$ if the relation $R$ is satisfied, and $\mathbf{I}(R) = 0$ otherwise. For a given value $\rho = \lambda/2\mu$, and a given threshold $x = n + r$ where $0 \le r < 1$, the steady state probabilities were computed as the asymptotic solution of the following transition equations:

$$\begin{aligned} p_{ij}(t + dt) = {} & p_{ij}(t)(1 - \lambda dt) + \mu p_{i,j+1}(t)dt \\ & - \mu p_{ij}(t)dt\mathbf{I}(i > 0) - \mu p_{ij}(t)dt\mathbf{I}(j > 0) \\ & + \mu p_{i+1,j}(t)dt\mathbf{I}(i \le n - 1, \text{ or } i = n \text{ and } r > 0) \\ & + \lambda p_{i-1,j}(t)dt\mathbf{I}(0 < i < n) \\ & + \lambda r p_{i-1,j}(t)dt\mathbf{I}(i = n + 1 \text{ and } r > 0) \\ & + (1 - r)\lambda p_{i,j-1}(t)dt\mathbf{I}(i = n \text{ and } j > 0) \\ & + \lambda p_{i,j-1}(t)dt\mathbf{I}(i = n + 1 \text{ and } j > 0). \end{aligned}$$

The solution depends only on the ratio $\rho = \lambda/2\mu$ and not on the individual values of $\lambda$ and $\mu$. For any given value of $\rho$ we searched for the $x$ value satisfying the equilibrium conditions.

For each equilibrium value $x$ we also computed the average rates $\lambda_1$ and $\lambda_2$ by which customers enter Q1 and Q2, respectively. We then computed the ratio $R = \lambda_1/\lambda_2$.[4] We found that this ratio is always greater than 1.

Figure 3 illustrates the equilibrium threshold $x$ (a solid line) and the ratio $R$ (a dashed line) for $\rho \in [0, 0.95]$. We see that $x$ is a nondecreasing function of $\rho$, with intervals in which it has a constant integer value. This behavior is expected in view of Figure 2: A change in $\rho$ amounts to a shift of the function $k(x)$ and the intersection with the identity function is obtained at any given integer in a nondegenerate interval of $\rho$. Figure 4 provides a closer look at the functions for $0.5 \le \rho$.

We also see from the figure that $R$ is not a monotone function of $\rho$. Its general trend is decreasing in $\rho$ and this is always the case in the intervals in which $x$ remains constant. Finally, as hinted by the title of this

[4] We computed $R$ from the state probabilities though it can also be computed, for any given $x$, directly from a set of recursive equations. Closed form formulas are more easily obtained for integer $x$ values. For example, for $x = 1$ $R = r_1 \equiv 1/2\rho$ and for $x = 2$ $R = r_1(1 + r_1)$.

note, $R > 1$ for all $\rho$ values, answering the question posed in the introduction.

## 6. Concluding Remarks

We saw that for all values of the system's utilization factor $\rho$ the average arrival rate into Q1 is greater than the rate into Q2. Clearly, if both queues can be observed the demand will be split equally, on average, between the servers. Thus, it pays to supply the queue size information to the customers. It is interesting to note that the demand will also be split equally if none of the two queues can be observed. However, this state is not an equilibrium with respect to the servers' behavior; each then has an incentive to reveal the size of his queue and increase the fraction of the demand directed to his facility. We thus expect to find observable queues whenever this is technically possible and costless.

In our application to the case of two facilities (gas stations) located on the same road, the one located first has higher profit. A natural outcome will be that the second station will try to attract demand by reducing prices or offering other benefits. Equilibrium prices can then be computed. We did not follow this direction and leave it for future research.

### References

Çinlar, E. and R. L. Disney, "Stream of Overflows from a Finite Queue," *Oper. Res.*, 15 (1967), 131–134.

van Doorn, E. A., "On the Overflow Process from a Finite Markovian Queue," *Performance Evaluation*, 4 (1984), 233–240.

Edelson, N. M. and K. Hildebrand, "Congestion Tolls for Poisson Queueing Processes," *Econometrica*, 43 (1975), 81–92.

Glazer, A., "The Strategy of Candidate Ambiguity," *American Political Sci. Review*, 84 (1990), 237–241.

—— and R. Hassin, "On the Economics of Subscriptions," *European Economic Review*, 19 (1982), 343–356.

Guérin, R. and L. Y. C. Lien, "Overflow Analysis for Finite Waiting Room Systems," *IEEE Trans. on Communications*, 38 (1990), 1569–1577.

Hassin, R., "Consumer Information in Markets with Random Product Quality: The Case of Queues and Balking," *Econometrica*, 54 (1986), 1185–1195.

—— and M. Haviv, "Equilibrium Threshold Strategies: The Case of Queues with Priorities," to appear in *Oper. Res.*

Haviv, M., "Stable Strategies for Processor Sharing Systems," *European J. Oper. Res.*, 13 (1991), 103–106.

Hlynka, M., D. A. Stanford, W. H. Poon, and T. Wang, "Observing Queues Before Joining," *Oper. Res.*, 42 (1994), 365–371.

Luski, I., "On Partial Equilibrium in a Queueing System with Two Servers," *Review of Economic Studies*, 43 (1976), 519–525.

Maynard Smith, J., *Evolution and Game Theory*, Cambridge University Press, Cambridge, MA, 1982.

Whitt, W., "Deciding Which Queue To Join: Some Counterexamples," *Oper. Res.*, 34 (1986), 55–62.

*Accepted by Linda V. Green; received August 29, 1994. This paper has been with the author $1\frac{1}{2}$ months for 2 revisions.*