

Equilibrium in Queues under Unknown Service Rates and Service Value

Laurens G. Debo	Senthil K. Veeraraghavan
<i>Chicago Booth School of Business</i>	<i>Wharton School</i>
<i>University of Chicago</i>	<i>University of Pennsylvania</i>
<i>Laurens.Debo@chicagobooth.edu</i>	<i>senthilv@wharton.upenn.edu</i>

January 2011

Abstract

We study a single queue joining equilibrium when there is uncertainty in the consumers' minds about the service rate and value. Without such uncertainty, the joining equilibria are characterized by means of a single threshold queue length above which consumers do not join (Naor, 1969). We show that in the presence of such uncertainty, the equilibrium joining strategy is not fully characterized by a single threshold. A “sputtering equilibrium” might exist. In the sputtering equilibrium, the queue length generally remains within a threshold, but reaches another, strictly higher, threshold, depending on the outcome of the randomized decision of the consumer arriving at the lower threshold. We discuss when and why sputtering equilibria exist.

Keywords: *Queueing games, Markov Perfect Bayesian Equilibrium, Threshold policies.*

1. Introduction

This note contributes to the stream of research on consumer queue joining behavior. Equilibrium queue joining behavior have been examined beginning with seminal work by Naor (1969). The reader is referred to Hassin and Haviv (2003) for an excellent overview. The established equilibrium result is as follows: When there are waiting costs, the consumers follow a *threshold* strategy. In other words, they join the queue if the length of the queue is below a threshold. Above the threshold, congestion effects dominate any value from service and consumers balk from the service. In the operations research literature, such negative externalities are extensively studied in a queueing context.

In this note, we relax two key assumptions in Naor's classical queue joining game: the observability of the service rates and their service values. We do so by modeling the service rate and value as a random variable whose realization is not observed by the consumers.

We demonstrate that unobservable service value and service rate may result in a “sputtering” equilibrium, where the queue length at the server generally resides within a lower threshold, but may *probabilistically* break through to a strictly higher threshold value, depending on the choice of the consumer who mixes between joining and balking at that particular state. This sputtering equilibrium fundamentally expands on the threshold equilibria seen in queues with rational decision-making consumers. A sputtering equilibrium may emerge when service value and service rate are negatively correlated. Thus, we find that the focus on simple threshold policies in queueing games with uncertainty about the service rate and value might be restrictive.

Non-threshold joining strategies have been identified in queuing games with priorities (Hassin and Haviv, 1996). Recent papers including Debo et al. (2010) and Veeraraghavan and Debo (2010) have been studying informational externalities associated with queue joining decisions under waiting costs. In both papers, non-threshold queue joining strategies emerge in queuing games with heterogeneously informed consumers. In our note, just as in Naor's model, consumers are homogenous and there are no priorities. Hence, we show that the non-threshold queue joining behavior emerges even in the canonical Naor (1969) model when the service rates and service values are not observable and negatively correlated. Thus, our analysis fundamentally expands on the queueing games literature with known service rates and values. Our analysis also suggests that in practice queue joining distributions cannot be generated by single threshold consumer joining strategies in the environments that we characterized. This may lead to further theoretical and empirical exploration.

2. Model

Consumers arrive sequentially to the server according to a Poisson process with parameter Λ . If the arriving consumers cannot be immediately served, they wait and form a queue. The queue discipline is first-come, first-served (FCFS). All consumers incur a disutility of $c > 0$ per unit time while waiting to complete the service.

The server belongs to one of the two types, $\omega \in \{h, \ell\}$. The utility that a consumer obtains from the server of type ω is v_ω , with $v_h \geq v_\ell$. The consumer's prior about the server's type is $\Pr(\omega = h) = p$. We address the server with service value v_h as the high-type or the high-quality server. (Thus the low-quality server has $v = v_\ell$.)

The service time of both types $\omega \in \{h, \ell\}$ is exponentially distributed. There are two types of processes, a *slow* process and a *fast* process. The mean service time of the slow (fast) process is $1/\underline{\mu}$ ($1/\bar{\mu}$) (i.e., $\underline{\mu} < \bar{\mu}$). As we have two server types and two types of service processes, we analyze four situations. For notational ease, we introduce $\boldsymbol{\mu} \doteq (\mu_h, \mu_\ell)$. Specifically, we use $\boldsymbol{\mu}_{ff} = (\bar{\mu}, \bar{\mu})$, $\boldsymbol{\mu}_{ss} = (\underline{\mu}, \underline{\mu})$, $\boldsymbol{\mu}_{sf} = (\underline{\mu}, \bar{\mu})$ and $\boldsymbol{\mu}_{fs} = (\bar{\mu}, \underline{\mu})$ as shorthand notations for the four situations. For any vector $\boldsymbol{\mu}_\sigma$, with $\sigma \in \{ss, ff, sf, fs\}$, the first letter refers to the service speed of the high-quality service (slow or fast), while the second letter refers to the speed of the low quality service (slow or fast). For instance, $\boldsymbol{\mu}_{sf} = (\underline{\mu}, \bar{\mu})$ corresponds to the situation when the high-quality server has slow service speed, and the low quality server has fast service speed. The consumer knows σ , but, does not know the realization of the server's type.

The model parameters are $(\Lambda, v_h, v_\ell, \bar{\mu}, \underline{\mu})$. Throughout this note, we assume that the value generated per unit of time by the high-quality server is always higher than that low-quality server's. That is, $v_h \underline{\mu} > v_\ell \bar{\mu}$.

The game proceeds as follows. First, Nature determines the server's type (with probability p , $\omega = h$). The server's type determines the service value (v_ω) and speed (μ_ω). Then, consumers arrive and observe the queue length, n , based on which they decide whether to join the queue or balk. The consumers maximize their expected net utility (i.e. service value minus the expected

waiting costs), conditional on their available information.

The Consumer Strategies and Beliefs: All consumers decide the probability of joining the service after observing the queue length $n \geq 0$. We denote this probability by $\alpha(n)$. The consumer's updated belief that the type of the server is high after observing a queue of length n is denoted by $\gamma(n)$. In short, $\boldsymbol{\alpha} = (\alpha(0), \alpha(1), \alpha(2), \dots)$ and $\boldsymbol{\gamma} = (\gamma(0), \gamma(1), \gamma(2), \dots)$ denote the consumer's joining strategy and the consumer's updated belief respectively. We need to determine the equilibrium strategies of all consumers: $\boldsymbol{\alpha}^*$ and $\boldsymbol{\gamma}^*$. We specify the equilibrium conditions next.

The Equilibrium Conditions: Consider a randomly arriving consumer. Suppose all other consumers are joining according to some strategy $\boldsymbol{\alpha}$. The consumer's expected utility of joining the queue, denoted by $u(n, \gamma)$, is a function of the queue length upon arrival, n , his/her belief about the server's type, γ . We now define:

Definition 1 (Equilibrium). *The strategies $\boldsymbol{\alpha}^*$, and beliefs $\boldsymbol{\gamma}^*$ form a Markov Perfect Bayesian Equilibrium (Maskin and Tirole 2001) if:*

(i)-the consumers are rational: for each $n \in \mathbb{N}$,

$$\alpha^*(n) \in \arg \max_{\alpha' \in [0,1]} \alpha' \times u(n, \gamma^*). \quad (1)$$

(ii)-beliefs are consistent: The belief $\gamma^*(n)$ satisfies Bayes' rule on all queue lengths that are reached with strictly positive probability in the long run under the strategy $\boldsymbol{\alpha}^*$.¹

The condition (i) of Definition 1 is referred to as the *rationality* condition for the consumers. The condition (ii) of Definition 1 is referred to as the *consistency* condition of the beliefs. Using the conditions, we can now analyze the queue joining equilibrium.

3. Analysis of Queue Joining Equilibrium

A consumer's utility can be written as,

$$u(n, \gamma) = \gamma(n) v_h + (1 - \gamma(n)) v_\ell - c(n+1) (\gamma(n) / \mu_h + (1 - \gamma(n)) / \mu_\ell). \quad (2)$$

With Equation (2), we can rewrite condition (i) of Definition 1 for consumers in terms of the likelihood ratio of the updated probability that the service is of high quality when a consumer observes queue length n (to updated probability that the service is of low quality). Now, instead of the updated belief, $\gamma(n)$, it will be convenient to introduce the likelihood ratio, $l(n)$, satisfying: $\gamma(n) / (1 - \gamma(n)) = l_0 \times l(n)$, where $l_0 = p / (1 - p)$. In other words, we decompose $\gamma(n)$ into the consumer's *prior likelihood* (that the server is of high quality), l_0 , and the *likelihood ratio* of observing n consumers upon arrival, $l(n)$. Then, with Equation (2), we obtain the following

¹When a queue length is not reached with positive probability, the belief is irrelevant, as well as the action at that queue length.

condition to make a consumer join a queue of length n :

$$u(n, \gamma) > 0 \Leftrightarrow l(n) > L(n), \text{ where } L(n) \doteq \frac{1}{l_0} \frac{-v_\ell + c(n+1)\frac{1}{\mu_\ell}}{v_h - c(n+1)\frac{1}{\mu_h}}. \quad (3)$$

L can be interpreted as the *minimum likelihood ratio* that is required to make a consumer join a queue of length n . The rationality condition (i) of Definition 1 determines $\alpha(n)$ for any n , which is 1 (0) when $l(n) > L(n)$ ($l(n) < L(n)$).

Now, we discuss condition (ii) of Definition 1. This condition needs to be imposed on $l(n)$. To that end, let $\pi(n, \alpha, \mu)$ be the long run probability that n consumers are in the system when the consumer strategy is α and the server's service rate is μ . For a given service rate and consumer joining strategy, the stochastic process that describes the queue length is a birth and death process, which allows us to characterize the long-run probability distribution. (See Ross 1996, pp. 254.)

Suppose consumers follow the strategy profile α . Then, for both server types $\omega \in \{h, \ell\}$, the stationary probability of observing a queue of length n , $\pi(n, \alpha, \mu)$, is:

$$\pi(n, \alpha, \mu) = \pi(0, \alpha, \mu) \prod_{j=0}^{n-1} \frac{\alpha(j)\Lambda}{\mu} \text{ where } \pi(0, \alpha, \mu) = \left[1 + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \frac{\alpha(j)\Lambda}{\mu} \right]^{-1}. \quad (4)$$

With the PASTA property (Wolff, 1982), $\pi(n, \alpha, \mu)$, is also the probability that a randomly arriving consumer observes n consumers in the system. The posterior probability that the server's quality is high after observing a queue length of n is $\frac{p\pi(n, \alpha, \mu_h)}{p\pi(n, \alpha, \mu_h) + (1-p)\pi(n, \alpha, \mu_\ell)}$. The posterior likelihood is then $l_0 \times \frac{\pi(n, \alpha, \mu_h)}{\pi(n, \alpha, \mu_\ell)}$. The consistency condition of Definition 1(ii) becomes: $l(n) = \frac{\pi(n, \alpha, \mu_h)}{\pi(n, \alpha, \mu_\ell)}$ for all queue lengths that are reached with positive probability in equilibrium.² With Equation (4), notice that $\frac{\pi(n, \alpha, \mu_h)}{\pi(n, \alpha, \mu_\ell)}$ becomes equal to $\frac{\pi(0, \alpha, \mu_h)}{\pi(0, \alpha, \mu_\ell)} \left(\frac{\mu_\ell}{\mu_h}\right)^n$. As the first factor is equal to $l(0)$, the consistency condition Definition 1(ii) becomes:

$$l(0) = \frac{\pi(0, \alpha, \mu_h)}{\pi(0, \alpha, \mu_\ell)} \text{ and } l(n) = l(0) \left(\frac{\mu_\ell}{\mu_h}\right)^n \text{ for } n \geq 1. \quad (5)$$

We exploit this structure to obtain the conditions for the equilibrium: instead of finding a fixed point α and γ that satisfies Definition 1(i) and (ii), we reduce the equilibrium characterization of Definition 1(i-ii) to a single-dimensional fixed point characterization in Proposition 1. This is because the likelihood ratio at the empty queue, $l(0)$, along with Equation (5) immediately determines $l(n)$ for any $n \geq 1$. With Equation (3), the rational actions can be immediately determined by comparing $l(n)$ with $L(n)$. Thus, the only unknown is $l(0)$.

In the following Proposition, we characterize the equilibrium by letting φ be a conjecture of the equilibrium $l(0)$. For any conjecture φ of $l(0)$, Equation (5) provides a conjecture of $l(n)$, which determines a rational strategy (i.e. join iff $l(n) > L(n)$, or randomize when $l(n) = L(n)$, see Equation (3)). This rational structure is an equilibrium if the likelihood ratio it induces at the

² $l(n)$ is indeterminate whenever $\pi(n, \alpha, \mu_h)$ and $\pi(n, \alpha, \mu_\ell)$ are both zero.

empty queue (obtained from Equation 4) is equal to φ . We refer to this fixed point as $\hat{\varphi}$. In that case, with $\hat{\varphi}$, a strategy profile, α^* , that is rational (Definition 1(i) is satisfied) and a corresponding belief, γ^* , that is consistent (Definition 1(ii) is satisfied for all $n \geq 0$) is constructed as described below:

Proposition 1. *Let $\hat{n} = (\hat{n}_0, \hat{\alpha}_0, \hat{n}_1)$ where $\hat{n}_0 \in \mathbb{N}$, $\hat{\alpha}_0 \in [0, 1)$ and $\hat{n}_1 \in \mathbb{N}$ ($\hat{n}_1 > \hat{n}_0$) and*

$$\Psi(\hat{n}) = \frac{\sum_{k=0}^{\hat{n}_0} \left[\frac{\Lambda}{\mu_\ell} \right]^k + \hat{\alpha}_0 \sum_{k=\hat{n}_0+1}^{\hat{n}_1} \left[\frac{\Lambda}{\mu_\ell} \right]^k}{\sum_{k=0}^{\hat{n}_0} \left[\frac{\Lambda}{\mu_h} \right]^k + \hat{\alpha}_0 \sum_{k=\hat{n}_0+1}^{\hat{n}_1} \left[\frac{\Lambda}{\mu_h} \right]^k}, \quad (6)$$

then,

(i) *there exists a ‘classical’ threshold equilibrium (pure or mixed if) for some $\hat{\varphi} \in \mathbb{R}^+$, and $\hat{n}_1 = \hat{n}_0 + 1$ the following conditions are satisfied:*

$$\hat{\varphi} = \Psi(\hat{n}_0, \hat{\alpha}_0, \hat{n}_0 + 1) \quad (7)$$

and

$$\begin{cases} \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^n \geq L(n), & \text{for } 0 \leq n < \hat{n}_0, \\ \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^n < L(n), & \text{for } n \geq \hat{n}_0. \end{cases} \quad (8)$$

where the first inequality is binding when $\alpha_0 > 0$.

(ii) *there exists a non-threshold equilibrium if for some $\hat{\varphi} \in \mathbb{R}^+$, and $\hat{n}_1 > \hat{n}_0 + 1$ the following conditions are satisfied: the condition of Equation (7) and*

$$\begin{cases} \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^n > L(n), & \text{for } 0 \leq n < \hat{n}_0 \text{ (if } 0 < \hat{\alpha}_0), \\ \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^{\hat{n}_0} = L(\hat{n}_0), & \text{for } n = \hat{n}_0, \\ \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^n > L(n), & \text{for } \hat{n}_0 < n < \hat{n}_1, \\ \hat{\varphi} \left(\frac{\mu_\ell}{\mu_h} \right)^n < L(n), & \text{for } n \geq \hat{n}_1. \end{cases} \quad (9)$$

Proposition 1(i) provides conditions for a pure or mixed strategy consumer equilibria: When $\hat{\alpha}_0 = 0$, a pure strategy equilibrium exists in which consumers join at any queue length that is strictly less than the threshold \hat{n}_0 , and balk from queues equal to or longer than \hat{n}_0 . When $\hat{\alpha}_0 > 0$, and $\hat{n}_1 = \hat{n}_0 + 1$, consumers join at any queue length that is strictly less than the threshold \hat{n}_0 , join with probability $\hat{\alpha}_0$ at the queue length \hat{n}_0 , otherwise, they balk. In this case, $\alpha^* = (1, 1, \dots, \hat{\alpha}_0, 0, 0, \dots)$, where $\hat{\alpha}_0$ is on the $\hat{n}_0 + 1$ st position. γ^* follows immediately from the long run probability distributions that are characterized by Equation (4). This is an extension of a pure strategy equilibrium, as discussed in Hassin and Haviv (2003), p. 7-8. We refer to this equilibrium simply as the *classical threshold equilibrium*. Proposition 1(ii) indicates that this classical equilibrium strategy is not always an equilibrium. When $\hat{n}_1 > \hat{n}_0 + 1$, Proposition 1(ii) identifies a non-threshold equilibrium with mixing (with probability $\hat{\alpha}_0$) at some queue length (\hat{n}_0), while joining

at longer queues and balking at queue length \hat{n}_1 . In this case, $\alpha^* = (1, 1, \dots, \hat{\alpha}_0, 1, \dots, 1, 0, 0, \dots)$, where $\hat{\alpha}_0$ is on the $(\hat{n}_0 + 1)$ -st position and the last 1 is at the $(\hat{n}_1 + 1)$ -st position. Again, γ^* follows immediately from the long run probability distributions.

In the following subsections, we elaborate further the intuition behind Proposition 1. To that end, we introduce $\nu(\varphi)$, which is the set of solutions to the $\varphi(\mu_\ell/\mu_h)^\nu = L(\nu)$. This is the continuous version of Equation (8), determining \hat{n}_0 . Define now $\hat{n}(\varphi) \doteq \lfloor \min \nu(\varphi) \rfloor$. $\hat{n}(\varphi)$ is the lowest queue length at which the consumer has negative expected utility, and hence balks when the conjecture of the likelihood ratio at the empty queue is φ .

Keeping all else equal, a larger φ means a higher likelihood ratio at the empty queue. Of course only the equilibrium value of φ will be the equilibrium likelihood ratio. By means of Equation (8), a higher likelihood ratio at every queue length. Therefore, the balking threshold $\hat{n}(\varphi)$ is non-decreasing in φ . Now, define:

$$\Phi(\hat{n}) \doteq \Psi((\hat{n}, 0, \cdot)), \quad (10)$$

i.e. the likelihood ratio at the empty queue when the consumer joining strategy is a pure threshold strategy determined by balking at \hat{n} . As $\hat{n}(\varphi) \in \mathbb{N}$, the function $\Phi(\hat{n}(\varphi))$ is a (discontinuous) staircase function in φ . In the following subsections, $\Phi(\hat{n}(\varphi))$ will be discussed further. We also provide intuition for the equilibrium structure of Proposition 1.

3.1 Classical threshold equilibrium at \hat{n}_0 with $\hat{\alpha}_0 = 0$

A pure strategy threshold joining equilibrium (with $\hat{\alpha}_0 = 0$) exists if there is a solution, $\hat{\varphi}$, to:

$$\varphi = \Phi(\hat{n}(\varphi)). \quad (11)$$

In words, if the conjecture about the likelihood ratio at the empty queue, $\hat{\varphi}$, leads to a rational joining strategy with threshold $\hat{n}_0 = \hat{n}(\hat{\varphi})$, which in turn leads to a likelihood ratio at the empty queue, $\Phi(\hat{n}_0)$ that is consistent with (i.e. equal to) $\hat{\varphi}$, then the two equilibrium conditions of Definition 1(i) and (ii) are satisfied. \hat{n}_1 is irrelevant as it is reached with zero probability.

Assume now that at one of the discontinuous points of $\hat{n}(\varphi)$, say, $\hat{\varphi}$, the consumers' threshold changes from \hat{n}_0 to \hat{n}_1 and causes the likelihood ratio induced by the threshold to change from a level strictly below (or above) $\hat{\varphi}$ to a level strictly above (or below) $\hat{\varphi}$. Formally, there exists a $\hat{\varphi}$ such that $\hat{n}(\hat{\varphi} - \epsilon) = \hat{n}_0$ and $\hat{n}(\hat{\varphi} + \epsilon) = \hat{n}_1$ for which $\Phi(\hat{n}_0) < \Phi(\hat{n}_1)$ (or: $\Phi(\hat{n}_0) > \Phi(\hat{n}_1)$). When $\hat{\varphi} \in (\Phi(\hat{n}_0), \Phi(\hat{n}_1))$ (or: $\hat{\varphi} \in (\Phi(\hat{n}_1), \Phi(\hat{n}_0))$), $\hat{\varphi}$ cannot characterize an pure strategy equilibrium. In that case, we can construct a consumer joining equilibrium involving mixing at \hat{n}_0 , i.e. $\alpha_0 > 0$.

We consider two cases involving mixing in the following two subsections. In the first case, we analyze $\hat{n}_1 = \hat{n}_0 + 1$. This corresponds with a classical threshold joining equilibrium with randomization. In the second case, $\hat{n}_1 > \hat{n}_0 + 1$. We will denote the latter as ‘sputtering’ equilibrium, because at \hat{n}_0 the queue ‘sputters’ before increasing to \hat{n}_1 due to the mixing probability $\hat{\alpha}_0$ at \hat{n}_0 .

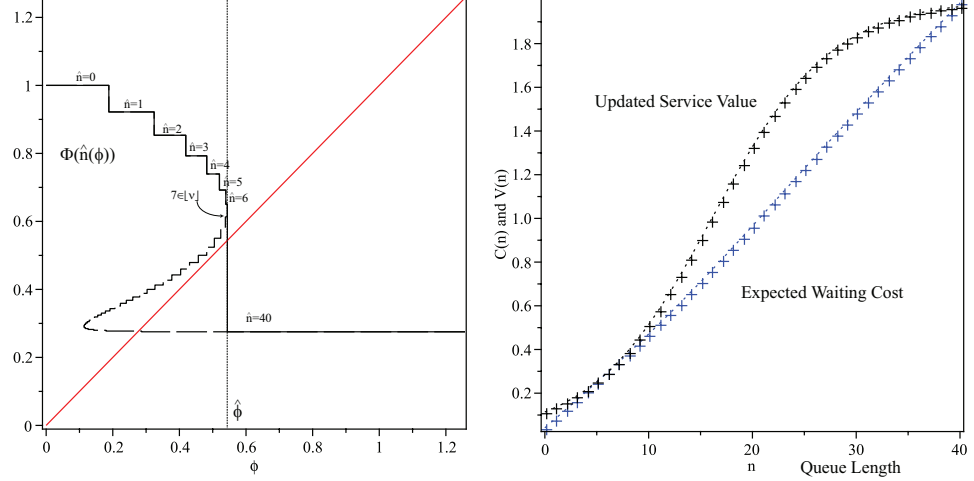


Figure 1: *Determination of equilibrium for $\mu = \mu_{sf}$; $\Phi(\hat{n}(\varphi)) = \phi$ (left panel) and the Updated Service Value, $\hat{\gamma}(n)v_h + (1 - \hat{\gamma}(n))v_\ell$, vs. the Updated Waiting Cost, $\hat{\gamma}(n)c(n+1)/\mu_h + (1 - \hat{\gamma}(n))c(n+1)/\mu_\ell$ (right panel). Demonstration of the sputtering equilibrium with randomization at $n = 6$ and a balking threshold at $n = 40$.*

3.2 Classical threshold equilibrium at \hat{n}_0 with randomization $\hat{\alpha}_0 > 0$ and $\hat{n}_1 = \hat{n}_0 + 1$

When $\hat{n}_1 = \hat{n}_0 + 1$, it is easy to see from the definitions of Equations (6) and (10) that:

$\Psi((\hat{n}_0, 0, \hat{n}_0 + 1)) = \Phi(\hat{n}_0)$ and $\Psi((\hat{n}_0, 1, \hat{n}_0 + 1)) = \Phi(\hat{n}_0 + 1)$. Then, by continuity of $\Psi((\hat{n}_0, \alpha, \hat{n}_0 + 1))$ in α , *any* likelihood ratio in $[\Phi(\hat{n}_0), \Phi(\hat{n}_0 + 1)]$ is achieved with a mixed strategy, $\alpha \in [0, 1]$ at \hat{n}_0 .

Thus, there exists an equilibrium with mixing at \hat{n}_0 such that the likelihood ratio of that strategy is equal to $\hat{\varphi}$. Both conditions (i) and (ii) of Definition 1 are satisfied. It follows that a classical mixed strategy extension of a pure strategy threshold (Hassin and Haviv, 2003) is an equilibrium; i.e. all consumers join queues that are strictly shorter than \hat{n}_0 , randomize with probability $\hat{\alpha}_0$ at a queue of length \hat{n}_0 , and balk at queues that are longer than or equal to $\hat{n}_0 + 1$.

3.3 Non-threshold or ‘Sputtering’ equilibrium with $\hat{\alpha}_0 > 0$ and $\hat{n}_1 > \hat{n}_0 + 1$

When $\hat{n}_1 > \hat{n}_0 + 1$, an extension of a pure threshold strategy with mixing with probability α_0 at the threshold \hat{n}_0 and balking at $\hat{n}_0 + 1$ cannot cover any likelihood ratio in $[\Phi(\hat{n}_0), \Phi(\hat{n}_1)]$ when $\Phi(\hat{n}_0 + 1) < \Phi(\hat{n}_1)$ and $\hat{\varphi} \in (\Phi(\hat{n}_0 + 1), \Phi(\hat{n}_1)]$. In that case, mixing at \hat{n}_0 only covers the likelihood ratio range of $[\Phi(\hat{n}_0), \Phi(\hat{n}_0 + 1)]$ and leaves the range $(\Phi(\hat{n}_0 + 1), \Phi(\hat{n}_1)]$ uncovered. If $\hat{\varphi}$ lies in the latter range, the classical mixed strategy extension cannot determine the equilibrium.

Proposition 1(ii) identifies a non-threshold strategy. Recall that Φ (Equation (10)) is a special case of Ψ (Equation (6)). By continuity of $\Psi((\hat{n}_0, \alpha, \hat{n}_1))$ in α , *any* likelihood ratio in $[\Phi(\hat{n}_0), \Phi(\hat{n}_1)]$ is achieved with a mixed strategy, $\alpha \in [0, 1]$ at \hat{n}_0 (as can be seen from the definitions). Hence, there exists a randomization probability $\hat{\alpha}_0$ that $\hat{\varphi}$ is reached if it lies in $(\Phi(\hat{n}_0 + 1), \Phi(\hat{n}_1)]$.

An Example of Sputtering Equilibrium: Now, we provide an illustrative example of a case

when no threshold consumer joining equilibrium exists. Let $v_h = 2$, $v_\ell = 0$, $p = 0.1$, $c = 0.051$ and $\underline{\mu} = 1.05$ and $\bar{\mu} = 1.25$. Let $\boldsymbol{\mu} = \boldsymbol{\mu}_{sf}$, that is: the high-quality server is slow while the low-quality server is fast. In Figure 1, left panel, we illustrate $\Phi(\hat{n}(\varphi))$. We indicate again the value of $\hat{n}(\varphi)$ on each horizontal segment where $\hat{n}(\varphi)$ is constant. Notice that at $\hat{\varphi} = 0.5439$, \hat{n} increases from $\hat{n}_0 = 6$ to $\hat{n}_1 = 40$. We obtain a non-threshold equilibrium with $\hat{\varphi} = 0.5439$, $\hat{n}_0 = 6$ and $\hat{\alpha}_0 = 0.1170$ and $\hat{n}_1 = 40$. This means that consumers always join the queue as long as its length is less than 6, they join with probability 0.1170 when the queue length is 6, they always join when the queue length is between 7 and 39, and they balk from any queue that is 40 or longer.

Now, we explain the intuition behind this result: Recall that $\nu(\varphi)$ is the set of real roots of $\varphi(\mu_\ell/\mu_h)^\nu = L(\nu)$. $\lfloor \nu(\varphi) \rfloor$ determines the queue length at which rational consumers do not join, assuming that φ is the likelihood ratio at the empty queue. On the left panel, we plot in dashed lines the correspondence $\Phi(\lfloor \nu(\varphi) \rfloor)$, where for certain values of φ , $\nu(\varphi)$ can take three different values. The values of $\lfloor \nu(\varphi) \rfloor$ over the dashed branch are between 7 and 40. Recall also that $\hat{n}(\varphi) = \lfloor \min \nu(\varphi) \rfloor$, i.e. the lowest queue length at which a consumer balks. Hence, whenever there are multiple solutions in $\nu(\varphi)$, the solution on the solid, top branch is selected. As a result, the range of queue lengths between 7 and 39 is excluded from $\hat{n}(\varphi)$, which ‘jumps’ from 6 to 40 (see Figure 1, left panel). As $\hat{\varphi} \in (\Phi(7), \Phi(40)]$, no classical threshold equilibrium can exist. This can also be observed from the left panel of Figure 1 as the 45 degree line has no intersection with $\Phi(\hat{n}(\varphi))$. Thus, with Proposition 1(ii), a non-threshold equilibrium is identified with randomization at a queue of length 6 and balking at a queue of length 40. The right panel of Figure 1 illustrates the equilibrium consumer updated utility (value and waiting cost). Indeed, there is one queue length (6), at which the consumer is indifferent between joining and balking, while at queue length 40, the consumer balks.

Queue Dynamics: It is interesting to observe that a non-threshold equilibrium causes the following queue dynamics. As at \hat{n}_0 , consumers join the queue with a probability of less than one, thus it will be difficult for the queue to grow beyond \hat{n}_0 . But, once the queue is larger than \hat{n}_0 , all consumers join again with probability one. Hence, the server will observe that the queue ‘stalls’ from time to time at \hat{n}_0 . This is why we label this equilibrium as a ‘sputtering’ equilibrium. These queue dynamics are an immediate result of the non-threshold consumer strategy. In the following subsection, we discuss when this phenomenon occurs.

3.4 Equilibrium Conditions

In this section, we discuss some properties of the model. First, we derive a sufficient condition for the classical threshold consumer joining equilibrium (Lemma 2). Next, we discuss the number of possible equilibria (Lemma 3).

From the discussion in the previous section, it follows that it is important to assess when $\hat{n}(\varphi)$ increases in steps of +1, and when it has larger jumps. When $(\mu_h/\mu_\ell)^\nu L(\nu)$ is monotone in ν and $\nu(\varphi)$ is a singleton, then, obviously, $\hat{n}(\varphi)$ increases in steps of +1. In general, however, $(\mu_h/\mu_\ell)^\nu L(\nu)$ is not always monotone in ν .

$\nu(\varphi)$ can contain three solutions, as illustrated in Figure 1, left panel, where for a range of values of φ two solutions are indicated by means of a dashed line and one solution (the lowest value) by means of the solid line. The reason why there are three possible solutions could be understood from the right panel of Figure 1: As the service value is bounded between v_ℓ and v_h , the updated service value after observing the queue length follows an S-shape for this particular example: the updated value is bounded by v_h and v_ℓ and at very low and very high queue lengths, finding one extra consumer in the queue provides low additional evidence that the server is of high quality. The updated waiting cost, on the other hand, is monotone increasing. As a result, there may be three intersection points between the S-shaped updated value and the monotone increasing updated cost. Hence, $\hat{n}(\varphi)$ can increase in steps that are larger than +1 (in Figure 1, from 6 to 40 at $\hat{\varphi} = 0.5439$). The following Lemma provides a sufficient condition for $\hat{n}(\varphi)$ to be increasing in steps of +1:

Lemma 2. *When $C(\boldsymbol{\mu}) > 0$, where*

$$C(\boldsymbol{\mu}) = c + \frac{1}{4}(\mu_h v_h - \mu_\ell v_\ell) \ln\left(\frac{\mu_h}{\mu_\ell}\right),$$

then, $\hat{n}(\varphi)$ increases in steps of +1, otherwise, $\hat{n}(\varphi)$ may increase in steps that are larger than +1.

It follows from Lemma 2 that $C(\boldsymbol{\mu}_{fs}) = c + \frac{1}{4}(\bar{\mu}v_h - \underline{\mu}v_\ell) \ln\left(\frac{\bar{\mu}}{\underline{\mu}}\right)$ is strictly positive (as $\bar{\mu}v_h > \underline{\mu}v_\ell$ and $\bar{\mu}/\underline{\mu} > 1$) and, trivially, $C(\boldsymbol{\mu}_{ss}) = C(\boldsymbol{\mu}_{ff}) = c$ is strictly positive. This implies that when the server's service rate is independent of the server's quality, or, when the highest quality server is also the fastest server, threshold joining strategies always exist as the updated service value is either independent of the queue length, or decreases in queue length, while the updated service cost increases in the queue length. Only when the highest quality server is also the slowest and both service value and cost increase in queue length (that is: service value and rate are negatively correlated), a sputtering equilibrium may exist. As a consequence, if $C(\boldsymbol{\mu}_{sf}) = c + \frac{1}{4}(\underline{\mu}v_h - \bar{\mu}v_\ell) \ln\left(\frac{\underline{\mu}}{\bar{\mu}}\right) > 0$, a classical threshold strategy is a consumer joining equilibrium for any pure service rate strategy, $\boldsymbol{\mu}_\sigma$ with $\sigma \in \{ss, ff, sf, fs\}$. In terms of the primitives of our model: If the difference between the slow ($\underline{\mu}$) and the fast speed ($\bar{\mu}$) is not high, or, that the waiting costs, c , are high enough, provided that the high quality server operating at a slow speed creates more value per unit of service time than the low quality server operating at a fast service speed ($\underline{\mu}v_h > \bar{\mu}v_\ell$), the consumer joining equilibrium is a classical threshold strategy for any service rate strategy. Otherwise, a sputtering queue joining equilibrium may exist.

An Illustration of $C(\boldsymbol{\mu})$: The condition $C(\boldsymbol{\mu}) = 0$ is illustrated in Figure 2. The dashed lines are $\mu_h v_h = \mu_\ell v_\ell$ and $\mu_h = \mu_\ell$. Notice that $C(\boldsymbol{\mu}) < 0$ only when $\mu_h v_h > \mu_\ell v_\ell$ and $\mu_h < \mu_\ell$, i.e., for a given service rate of the low quality server; when the service rate of the high quality server is lower, but, not too low, it is possible that $C(\boldsymbol{\mu}) < 0$. $C(\boldsymbol{\mu}) > 0$ is a sufficient condition for a threshold equilibrium strategy. When $C(\boldsymbol{\mu}) > 0$, we can analyze the properties of threshold strategy equilibria, i.e. when $\hat{\alpha}_0 = 0$ or $\hat{\alpha}_0 > 0$ and $\hat{n}_1 = \hat{n}_0 + 1$. In Lemma 3, we characterize the number of pure strategy joining equilibria:

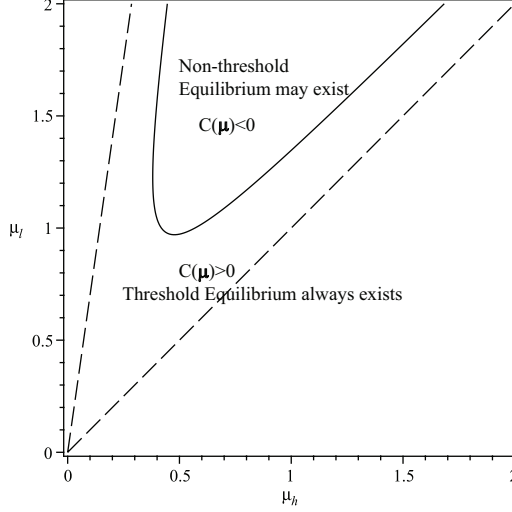


Figure 2: $C(\mu) = 0$ for $v_h = 1.75$, $v_\ell = 0.25$ and $c = 0.1051$. The dashed lines are $\mu_h v_h = \mu_\ell v_\ell$ and $\mu_h = \mu_\ell$.

Lemma 3. When $C(\mu) > 0$, then:

- (i) When $\mu_h < \mu_\ell$, there exists at most one pure strategy threshold equilibrium. If no pure strategy equilibrium exists, a mixed strategy equilibrium exists.
- (ii) When $\mu_\ell \leq \mu_h$, there exists at least one pure strategy threshold equilibrium. If more than one pure strategy equilibrium exists, a mixed strategy equilibrium exists.

The intuition behind Lemma 3 is the following: Consider the situation when the high quality server is the slowest server ($\sigma = sf$ or $\mu_h < \mu_\ell$); it is thus easily understood that the likelihood ratio at the empty queue, $\Phi(\hat{n})$, decreases in the joining threshold, \hat{n} .

Assume that \hat{n} is very low. The queue is mostly short. Hence, the recurrent state space (queue lengths) is small such that the probability of observing an empty queue when the product quality is high is not much different from the probability of observing an empty queue when the service quality is low. Hence, Φ is comparable to one.

When \hat{n} is high, there is enough variation in the queue lengths such that there is a difference in probability of observing an empty queue. In fact, the probability of observing an empty queue when the quality is high will be lower than the probability of observing an empty queue when the quality is low. Hence, Φ is less than one, and is thus decreasing in \hat{n} . As $\hat{n}(\varphi)$ is always increasing in φ , it follows that $\Phi(\hat{n}(\varphi))$ decreases in φ . As a result, there can be at most one point of intersection for $\Phi(\hat{n}(\varphi))$ and the 45 degree line, and, therefore, at most one pure strategy equilibrium. For such pure strategy equilibrium, $\hat{\varphi} < 1$. When the low quality server is slow ($\sigma = fs$), the opposite is true: $\Phi(\hat{n})$ increases. As a result, there can be multiple points of intersection for $\Phi(\hat{n}(\varphi))$ and the 45 degree line, and thus multiple pure strategy equilibria with $\hat{\varphi} > 1$ may exist.

Recall from Section 3 that $\hat{\varphi}$ is the likelihood ratio at the empty queue, $l(0)$, and we decomposed the posterior belief that the quality is high via $\gamma(n) / (1 - \gamma(n)) = l_0 \times l(n)$. Hence, in any

equilibrium in which the high-quality server is the slowest ($\sigma = sf$), as the equilibrium $l(0)$ is less than one, an empty queue is ‘bad news’ about the service quality (that is: $\gamma(0) < p_0$). Longer queues, however, make the consumer more confident that the quality is high (see Equation (9)). When the low quality server is slow ($\sigma = fs$), the opposite is true: empty queues carry good news about the quality ($\gamma(0) > p_0$) and longer queues make the consumer less confident that the quality is high.

4. Conclusions

In summary, combining Proposition 1 and Lemmas 2 and 3 yield the following conclusions.

1. When the high quality server is slow and the low quality server is fast, the updated value from the service increases in the queue length. However, long queues also imply long waiting times. Hence, both costs and value increase in queue length. When the difference between the service rates of the high- and the low-quality server is small, there will exist a threshold, at which the additional updated value does not compensate for the increase in the waiting costs. Consumers balk from the queue at this threshold.

2. When the difference in service rates between different services is large enough, a “sputtering” equilibrium involving randomization at some lower threshold emerges. Consumers are indifferent between joining and balking at that threshold, while, at strictly longer and shorter queue lengths, the consumers may strictly prefer joining the queue. Of course, as waiting costs grow without bound in the queue length and the service value is finite, there will exist a second, higher threshold above which no consumer ever joins. As a result, the queue generally resides at low lengths, and with some probability (depending on the random decision of one consumer), grows to a larger queue, since every one else joins at higher queue lengths.

To summarize, we find that the presence of informational uncertainty on service value and service rate can introduce a sputtering queue joining equilibrium. Thus, a focus on simple threshold policies in queuing games with uncertainty about the service rate and value might be restrictive.

References

- Debo, L., C. Parlour and U. Rajan, 2010. Inferring Quality from a Queue. Chicago Booth Working paper.
- Hassin, R. and M. Haviv. 2003. To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems. Kluwer Academic Publishers.
- Hassin, R. and M. Haviv. 1997. Equilibrium Threshold Strategies: The Case of Queues with Priorities. *Operations Research*. 45(6), pp. 966-973.
- Maskin, E and J. Tirole. 2001. Markov Perfect Equilibrium. *J. of Econ. Theory*. 100, pp. 191–219.
- Naor, P. 1969. The Regulation of Queue Size by Levying Tolls. *Econometrica*, 37, pp. 15-34.
- Ross, S. 1996. Stochastic Processes. Wiley and Sons, New York.

Veeraraghavan, S. and L. Debo. 2010. Herding in Queues under Waiting Costs: Rationality and Regret. Forthcoming in *M&SOM*.

Wolff, R. 1982. Poisson arrivals see time averages. *Operations Research*, 30, pp. 223-231.

Proofs

Proof of Proposition 1: We impose conditions (i) and (ii) of Definition 1 for a given joining profile α . We consider only strategy profiles of the following special form, parameterized by some $\mathbf{n} = (n_0, \alpha_0, n_1)$: $\alpha(j) = 1$ for $0 \leq j < n_0$ and $\alpha(n_0) = \alpha_0$ and $\alpha(j) = 1$ for $n_0 + 1 \leq j < n_1$ and $\alpha(j) = 0$ for $j \geq n_1 + 1$. Then, we can rewrite with Equation (4) $l(n)$ as:

$$l(0) = \Psi(\mathbf{n}) \text{ and } l(n) = l(0) \left(\frac{\mu_\ell}{\mu_h} \right)^n$$

(for all n that are reached with positive probability on the long run). With Equations (3) and (5), we can rewrite conditions (i) and (ii) of Definition 1 as:

(i)-consumers are rational when for each for all n that are reached with positive probability on the long run, they join when

$$\Psi(\mathbf{n}) \left(\frac{\mu_\ell}{\mu_h} \right)^n > (<) L(n) \Rightarrow \hat{\alpha}(n) = 1(0) \text{ and } \Psi(\mathbf{n}) \left(\frac{\mu_\ell}{\mu_h} \right)^n = L(n) \Rightarrow \hat{\alpha}(n) \in [0, 1]. \quad (12)$$

(ii)-beliefs are consistent: The belief $\hat{\gamma}(n)$ satisfies Bayes' rule when:

$$\hat{\gamma}(n) = \frac{\Psi(\mathbf{n}) \left(\frac{\mu_\ell}{\mu_h} \right)^n}{\Psi(\mathbf{n}) \left(\frac{\mu_\ell}{\mu_h} \right)^n + 1} \text{ for } n \geq 0.$$

Hence, (1) the belief on the queue lengths that are reached with strictly positive probability on the long run, $\hat{\gamma}(n)$ is completely specified by \mathbf{n} and (2) for a given \mathbf{n} , the rationality conditions are completely determined. Equation (12) determines thus the equilibrium conditions $\hat{\mathbf{n}}$, which determine then $\hat{\alpha}$.

Now, we introduce the variable $\varphi > 0$. In equilibrium, $\hat{\varphi}$, will be equal to $\Psi(\hat{\mathbf{n}})$. We replace $\Psi(\mathbf{n})$ in Equation (12) by φ and for any $\varphi > 0$, we define n_0 is the *lowest* queue length at which the consumer balks:

$$\hat{n}(\varphi) = \min\{n \in \mathbb{N} : \varphi \left(\frac{\mu_\ell}{\mu_h} \right)^n \geq L(n), \text{ for } 0 \leq n < n_0 \text{ and } \varphi \left(\frac{\mu_\ell}{\mu_h} \right)^n < L(n), \text{ for } n \geq n_0\}.$$

Recall that we introduced

$$\Phi(\hat{n}(\varphi)) = \Psi((\hat{n}(\varphi), 0, \cdot)).$$

The \cdot indicates that $\Psi((\hat{n}_0, 0, \cdot))$ does not depend on n_1 . As $\hat{n}(\varphi) \in \mathbb{N}$, $\hat{n}(\varphi)$ is discontinuously increasing in φ . Hence, $\Phi(\hat{n}(\varphi))$ is a discontinuous function in φ . We can extend $\Phi(\hat{n}(\varphi))$ to a correspondence, $\hat{\Phi}(\hat{n}(\varphi))$, where at any discontinuous point φ' of $\hat{n}(\varphi)$, where $\hat{n}(\varphi - \epsilon) = n_0$ and $\hat{n}(\varphi + \epsilon) = n_1$ for an arbitrary small, but strictly positive ϵ , the image of the correspondence is the set $[\Phi(n_0), \Phi(n_1)]$ if $\Phi(n_0) < \Phi(n_1)$ or $[\Phi(n_1), \Phi(n_0)]$ otherwise.

Existence of a fixed point of $\hat{\Phi}(\hat{n}(\varphi)) = \varphi$. Notice that $\varphi = 0$: $\hat{n}(0) \geq 0$ for which $\hat{\Phi}(\hat{n}(0)) > 0$ and for $\varphi \rightarrow \infty$: $\hat{n}(\infty) \leq \lfloor \frac{v_h \mu_h}{c} \rfloor < \infty$ for which $\hat{\Phi}(\hat{n}(\infty)) < \infty$. It follows that the correspondence has at least one fixed point, $\hat{\varphi}$, in $(0, +\infty)$. Next, we characterize a fixed point.

Characterization of a fixed point of $\hat{\Phi}(\hat{n}(\varphi)) = \varphi$. We consider three cases:

Case 1. Assume that for some continuous point of $\hat{n}(\varphi)$, $\hat{\varphi}$, the corresponding $\hat{n}_0 = \hat{n}(\hat{\varphi})$ satisfies

$$\hat{\varphi} = \Phi(\hat{n}(\hat{\varphi}))$$

Then, it is easy to see that the strategy profile $\hat{\alpha}$, defined by $\hat{\alpha}(j) = 1$ for $0 \leq j < \hat{n}_0$ and $\hat{\alpha}(\hat{n}_0) = 0$ satisfies (i) of Definition 1, by construction of $\hat{n}(\varphi)$ (as \hat{n}_0 is the *lowest* queue length at which the consumer balks, the consumer joins for all queue lengths strictly lower than \hat{n}_0 , as is assumed in the special structure of α) and also (ii) of Definition 1 because $\hat{\varphi} = l(0)$. Hence, $\hat{\alpha}$ satisfies conditions (i) and (ii) of Definition 1.

Case 2. Assume that no continuous point of $\hat{n}(\varphi)$ exists for which the above fixed point equality is satisfied. As the correspondence $\hat{\Phi}(\hat{n}(\varphi))$, there must exist at least one discontinuous point, $\hat{\varphi}$, that satisfies $\hat{\Phi}(\hat{n}(\hat{\varphi})) = \hat{\varphi}$. Now, let $\hat{n}_0 = \hat{n}(\hat{\varphi} - \epsilon)$ and assume that $\hat{n}(\hat{\varphi} + \epsilon) = \hat{n}_0 + 1$ then: either

$$\Phi(\hat{n}_0) > \hat{\varphi} > \Phi(\hat{n}_0 + 1) \text{ or } \Phi(\hat{n}_0) < \hat{\varphi} < \Phi(\hat{n}_0 + 1) \quad (13)$$

Now consider $\alpha(j) = 1$ for $0 \leq j < n_0$ and $\alpha(n_0) = \alpha_0$ and $\alpha(j) = 0$ for $j \geq n_0 + 1$. Then, condition (ii) for consistent beliefs becomes:

$$\hat{\varphi} = \Psi((\hat{n}_0 + 1, \alpha_0, \hat{n}_0 + 1)).$$

From continuity of the right hand side of the above expression in α_0 , which ranges from $\Phi(\hat{n}_0)$ (for $\alpha_0 = 0$) to $\Phi(\hat{n}_0 + 1)$ (for $\alpha_0 = 1$) and by the inequalities (13) it follows that there must exist a $\hat{\alpha}_0$ such that the rationality condition at the empty queue needs to be satisfied: $\hat{\varphi} = \Psi((\hat{n}_0 + 1, \hat{\alpha}_0, \hat{n}_0 + 1))$. Then, it is easy to see that the strategy profile $\hat{\alpha}$, defined by $\hat{\alpha}(j) = 1$ for $0 \leq j < \hat{n}_0$, $\hat{\alpha}(\hat{n}_0) = \hat{\alpha}_0$ and $\hat{\alpha}(j) = 0$ for $j \geq \hat{n}_0 + 1$ satisfies (i) of Definition 1, by construction of $\hat{n}(\varphi)$ and also (ii) of Definition 1 because $\hat{\varphi} = l(0)$. Hence, $\hat{\alpha}$ satisfies conditions (i) and (ii) of Definition 1.

Case 3. Assume again that no continuous point of $\hat{n}(\varphi)$ exists for which the above fixed point equality is satisfied. Now, assume that $\hat{n}(\hat{\varphi} + \epsilon) = \hat{n}_1 > \hat{n}_0 + 1$. Then, either:

$$\Phi(\hat{n}_0) > \hat{\varphi} > \Phi(\hat{n}_1) \text{ or } \Phi(\hat{n}_0) < \hat{\varphi} < \Phi(\hat{n}_1). \quad (14)$$

Similarly as in Case 2., consider $\alpha(j) = 1$ for $0 \leq j < n_0$ and $\alpha(n_0) = \alpha_0$ and introduce $\alpha(j) = 1$ for $n_0 + 1 \leq j < n_1$ and $\alpha(j) = 0$ for $j \geq n_1 + 1$. Notice that by definition of Ψ (Equation 6) for $\hat{n}_1 = \hat{n}(\varphi + \epsilon) > \hat{n}_0 + 1$:

$$\Phi(\hat{n}_0) = \Psi((\hat{n}_0, 0, \hat{n}_1)) \text{ and } \Phi(\hat{n}_1) = \Psi((\hat{n}_0, 1, \hat{n}_1)).$$

From continuity of $\Psi((\hat{n}_0, \alpha_0, \hat{n}_1))$ in α_0 , and by the inequalities (14) it follows that there must exist a $\hat{\alpha}_0$ such that the rationality condition at the empty queue needs to be satisfied: $\hat{\varphi} = \Psi((\hat{n}_0, \hat{\alpha}_0, \hat{n}_1))$. Again, it is easy to see that the strategy profile $\hat{\alpha}$, defined by $\hat{\alpha}(j) = 1$ for $0 \leq j < n_0$ and $\hat{\alpha}(n_0) = \hat{\alpha}_0$ and $\hat{\alpha}(j) = 1$ for $n_0 + 1 \leq j < n_1$ and $\hat{\alpha}(j) = 0$ for $j \geq n_1 + 1$ satisfies (i) of Definition 1 (by construction of $\hat{n}(\varphi)$) and (ii) of Definition 1 (because $\hat{\varphi} = l(0)$). Hence, $\hat{\alpha}$ satisfies conditions (i) and (ii) of Definition 1.

Case 1, Case 2 and Case 3 result in Equations (6), (7), (8) and (9). \square

Proof of Lemma 2: Define:

$$\varphi(\nu) \doteq \begin{cases} \left(\frac{\mu_h}{\mu_\ell}\right)^\nu L(\nu) & \text{for } \nu \in [\bar{\nu}_\ell, \bar{\nu}_h - 1] \\ +\infty & \text{for } \nu \geq \bar{\nu}_h - 1, \end{cases}$$

Where $\bar{\nu}_\omega = v_\omega \mu_\omega / c$. $\varphi(n)$ is the likelihood ratio upon observing an empty queue that will make the consumer indifferent between joining a queue of length n and balking, assuming that at queue lengths 0 to n , all consumers join and informed consumers only join when the server is of high quality, i.e. when $\hat{n}(\varphi') = n'$, then $\varphi(n') = \varphi'$. When $\varphi(n)$ is increasing in n , this means that as φ increases, $\hat{n}(\varphi)$ increases in increments of 1. When $\varphi(n)$ is not monotone, this means that $\hat{n}(\varphi)$ increases in increments of potentially more than 1. In order to establish this, we consider ν as a continuous variable and derive $\varphi(\nu)$. We find that over $[\bar{\nu}_\ell, \bar{\nu}_h - 1)$, the derivative can have zero or two zero points. We ignore the knife-edge case of one degenerate zero point. Let $\underline{\nu}$ and $\bar{\nu}$ be the zero points of $\frac{d}{d\nu}\varphi(\nu)$. It can easily be established that $\frac{d}{d\nu}\varphi(0) > 0$, hence, $\varphi(\nu)$ increases first, reaches a local maximum and then decreases to a local minimum, after which it increases again. It can also be established easily that $\lim_{\nu \rightarrow \bar{\nu}_h - 1} \varphi(\nu) = +\infty$. As a result, over $[\underline{\nu}, \bar{\nu}]$, where $\varphi(\nu)$ decreases, $\hat{n}(\varphi)$ may increase in jumps that are larger than 1. Now, we take the derivative:

$$\frac{d}{d\nu}\varphi(\nu) = \left\{ \frac{v_\ell - v_h}{\left(v_h - c \frac{\nu+1}{\mu_h}\right)^2} \frac{c}{\mu_h} + \frac{-v_\ell + c \frac{\nu+1}{\mu_\ell}}{v_h - c \frac{\nu+1}{\mu_\ell}} \ln\left(\frac{\mu_h}{\mu_\ell}\right) \right\} \frac{1}{l_0} \left(\frac{\mu_h}{\mu_\ell}\right)^\nu$$

Hence, it follows that

$$\begin{aligned} \frac{d}{d\nu}\varphi(\nu) &> 0 \Leftrightarrow \\ \frac{c \frac{1}{\mu_\ell} \left(v_h - c \frac{\nu+1}{\mu_h}\right) + \left(-v_\ell + c \frac{\nu+1}{\mu_\ell}\right) \left(c \frac{1}{\mu_h}\right)}{\left(v_h - c \frac{\nu+1}{\mu_h}\right)^2} + \frac{-v_\ell + c \frac{\nu+1}{\mu_\ell}}{v_h - c \frac{\nu+1}{\mu_\ell}} \left(\ln\left(\frac{\mu_h}{\mu_\ell}\right)\right) &> 0 \Leftrightarrow \\ \frac{c}{\mu_\ell} + \frac{\left(-v_\ell + c \frac{\nu+1}{\mu_\ell}\right)}{\left(v_h - c \frac{\nu+1}{\mu_h}\right)} \frac{c}{\mu_h} + \left(-v_\ell + c \frac{\nu+1}{\mu_\ell}\right) \ln\left(\frac{\mu_h}{\mu_\ell}\right) &> 0 \Leftrightarrow \\ c \frac{\mu_h v_h - \mu_\ell v_\ell}{\mu_h \mu_\ell} + \left(c \frac{\nu+1}{\mu_\ell} - v_\ell\right) \left(v_h - c \frac{\nu+1}{\mu_h}\right) \ln\left(\frac{\mu_h}{\mu_\ell}\right) &> 0 \\ \Rightarrow -\ln\left(\frac{\mu_h}{\mu_\ell}\right) \underbrace{\left(c \frac{\nu+1}{\mu_\ell} - v_\ell\right)}_{\geq 0} \underbrace{\left(v_h - c \frac{\nu+1}{\mu_h}\right)}_{\geq 0} &< c \frac{\mu_h v_h - \mu_\ell v_\ell}{\mu_h \mu_\ell}. \end{aligned} \quad (15)$$

This is a quadratic equation in ν . Hence, $\frac{d}{d\nu}\varphi(\nu) = 0$ has zero, or two real solutions, depending on the parameters. Notice that the left hand side reaches a maximum when

$$\begin{aligned} \frac{d}{d\nu} \left(\left(-v_\ell + c \frac{\nu+1}{\mu_\ell}\right) \left(v_h - c \frac{\nu+1}{\mu_h}\right) \right) &= 0 \Leftrightarrow \nu^* = -1 + \frac{\mu_h v_h + \mu_\ell v_\ell}{2c} \\ v_\ell - c \frac{\nu^* + 1}{\mu_\ell} &< 0 < v_h - c \frac{\nu^* + 1}{\mu_h} \Leftrightarrow -\frac{1}{2} \frac{\mu_h v_h - \mu_\ell v_\ell}{\mu_\ell} < 0 < \frac{1}{2} \frac{\mu_h v_h - \mu_\ell v_\ell}{\mu_h} \end{aligned}$$

and as a result, assuming that $\mu_h v_h - \mu_\ell v_\ell > 0$ is always satisfied because in the worst case, $\mu v_h > \bar{\mu} v_\ell$ (all other cases are trivially satisfied), ν^* lies in between $\bar{\nu}_\ell$ and $\bar{\nu}_h - 1$. Thus, we can write:

$$\left(-v_\ell + c \frac{\nu+1}{\mu_\ell}\right) \left(v_h - c \frac{\nu+1}{\mu_h}\right) \leq \frac{1}{4} \frac{(\mu_h v_h - \mu_\ell v_\ell)^2}{\mu_h \mu_\ell} \quad (16)$$

Plugging Equation (16) in Equation (15), we obtain that:

$$c \frac{\mu_h v_h - \mu_\ell v_\ell}{\mu_h \mu_\ell} > -\frac{1}{4} \frac{(\mu_h v_h - \mu_\ell v_\ell)^2}{\mu_h \mu_\ell} \ln \left(\frac{\mu_h}{\mu_\ell} \right) \Rightarrow \frac{d}{d\nu} \varphi(\nu) > 0, \forall \nu \in [\bar{\nu}_\ell, \bar{\nu}_h - 1]$$

or

$$C(\boldsymbol{\mu}) = c + \frac{1}{4} (\mu_h v_h - \mu_\ell v_\ell) \ln \left(\frac{\mu_h}{\mu_\ell} \right) > 0 \Rightarrow \frac{d}{d\nu} \varphi(\nu) > 0, \forall \nu \in [\bar{\nu}_\ell, \bar{\nu}_h - 1].$$

We have thus proven that when $C(\boldsymbol{\mu}) > 0$, $\varphi(\nu)$ is increasing. Hence $\hat{n}(\varphi)$ increases in jumps of one. When $C(\boldsymbol{\mu}) < 0$, there exist two roots of Equation (15), $\underline{\nu}$ and $\bar{\nu}$. $\varphi(\nu)$ is decreasing in between the two roots, which implies that $\hat{n}(\varphi)$ will never take values over $(\underline{\nu}, \bar{\nu})$. Hence $\hat{n}(\varphi)$ may increase in jumps that are larger than one. Notice that when $\frac{\mu_h}{\mu_\ell} > 1$, $C(\boldsymbol{\mu}) > 0$. Only when $\frac{\mu_h}{\mu_\ell} < 1$, we may obtain that $C(\boldsymbol{\mu}) < 0$. \square

Proof of Lemma 3: First, we write $\Phi(\nu)$ as follows:

$$\Phi(\nu) = \frac{\sum_{n=0}^{\nu} \left[\frac{\Lambda}{\mu_\ell} \right]^n}{\sum_{n=0}^{\nu} \left[\frac{\Lambda}{\mu_h} \right]^n} = \left(\frac{1 - \left[\frac{\Lambda}{\mu_\ell} \right]^{\nu+1}}{1 - \frac{\Lambda}{\mu_\ell}} \right) / \left(\frac{1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1}}{1 - \frac{\Lambda}{\mu_h}} \right)$$

We show that $\Phi(\nu)$ increases in ν iff $\mu_\ell < \mu_h$. Now, we consider ν as a continuous variable and determine the condition when $\Phi(\nu)$ is increasing:

$$\frac{1 - \frac{\Lambda}{\mu_h} \frac{d}{d\nu} \left(1 - \left[\frac{\Lambda}{\mu_\ell} \right]^{\nu+1} \right)}{1 - \frac{\Lambda}{\mu_\ell} \frac{d}{d\nu} \left(1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1} \right)} > \Phi(\nu) \Leftrightarrow \frac{1 - \frac{\Lambda}{\mu_h} \frac{d}{d\nu} \left(\left[\frac{\Lambda}{\mu_\ell} \right]^{\nu+1} \right)}{1 - \frac{\Lambda}{\mu_\ell} \frac{d}{d\nu} \left(\left[\frac{\Lambda}{\mu_h} \right]^{\nu+1} \right)} > \Phi(\nu)$$

as $\frac{d}{d\nu} (a^{\nu+1}) = a^{\nu+1} \ln a$, we obtain that

$$\begin{aligned} \frac{d}{d\nu} \Phi(\nu) > 0 &\Leftrightarrow \left(\frac{\mu_h}{\mu_\ell} \right)^{\nu+1} \frac{1 - \frac{\Lambda}{\mu_h} \ln \left(\frac{\Lambda}{\mu_\ell} \right)}{1 - \frac{\Lambda}{\mu_\ell} \ln \left(\frac{\Lambda}{\mu_h} \right)} > \frac{1 - \frac{\Lambda}{\mu_h} \frac{1 - \left[\frac{\Lambda}{\mu_\ell} \right]^{\nu+1}}{1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1}}}{1 - \frac{\Lambda}{\mu_\ell} \frac{1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1}}{1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1}}} \text{ or:} \\ \frac{d}{d\nu} \Phi(\nu) > 0 &\Leftrightarrow \left(\frac{\Lambda}{\mu_\ell} \right)^{\nu+1} \frac{\ln \left(\frac{\Lambda}{\mu_\ell} \right)}{1 - \left[\frac{\Lambda}{\mu_\ell} \right]^{\nu+1}} > \left(\frac{\Lambda}{\mu_h} \right)^{\nu+1} \frac{\ln \left(\frac{\Lambda}{\mu_h} \right)}{1 - \left[\frac{\Lambda}{\mu_h} \right]^{\nu+1}} \end{aligned} \quad (17)$$

Now, consider

$$x^{\nu+1} \frac{\ln(x)}{1 - x^{\nu+1}}$$

and let $z = x^{\nu+1}$, then notice

$$z \frac{\ln \left(z^{\frac{1}{\nu+1}} \right)}{1 - z} = \frac{1}{\nu+1} z \frac{\ln(z)}{1 - z}$$

and as $z \frac{\ln(z)}{1-z} < 0$ and continuous decreasing for $z > 0$. Hence, the condition of Equation (17) is satisfied iff $\frac{\Lambda}{\mu_\ell} > \frac{\Lambda}{\mu_h}$ or $\mu_h > \mu_\ell$, then $\frac{d}{d\nu} \Phi(\nu) > 0$, otherwise $\frac{d}{d\nu} \Phi(\nu) < 0$. \square