



Equilibrium in a finite capacity $M/M/1$ queue with unknown service rates consisting of strategic and non-strategic customers

S. Srivatsa Srinivas¹ · Rahul R. Marathe¹

Received: 29 September 2019 / Revised: 26 September 2020 / Accepted: 29 September 2020 /
Published online: 14 October 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

We consider an $M/M/1/\bar{N}$ observable non-customer-intensive service queueing system with unknown service rates consisting of strategic impatient customers who make balking decisions and non-strategic patient customers who do not make any decision. In the queueing game amongst the impatient customers, we show that there exists at least one pure threshold strategy equilibrium in the presence of patient customers. As multiple pure threshold strategy equilibria exist in certain cases, we consider the minimal pure threshold strategy equilibrium in our sensitivity analysis. We find that the likelihood ratio of a fast server to a slow server in an empty queue is monotonically decreasing in the proportion of impatient customers and monotonically increasing in the waiting area capacity. Further, we find that the minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers and non-decreasing in the waiting area capacity. We also show that at least one pure threshold strategy equilibrium exists when the waiting area capacity is infinite.

Keywords Queueing game · Service operations · Strategic behavior · Threshold strategy equilibrium

Mathematics Subject Classification 60K25 · 90B22 · 91A40

1 Introduction

We consider a finite capacity observable $M/M/1$ queueing system, which provides non-customer-intensive service and follows a first-in-first-out (FIFO) queue discipline.

✉ Rahul R. Marathe
rrmarathe@iitm.ac.in

S. Srivatsa Srinivas
srivats.sss@gmail.com

¹ Department of Management Studies, Indian Institute of Technology Madras, Chennai, India

In contrast to customer-intensive services where service value increases with service time, service value decreases with service time in non-customer-intensive services. The arriving customers are of two types—(1) impatient customers who are strategic and make balking decisions, (2) patient customers who are non-strategic, always join the queue, and do not make any decision at all. Nature first probabilistically selects the service rate (fast or slow). The impatient customers then update their beliefs on the service rate by observing the queue and then make balking decisions based on the net utility from the service.

A classic example of our setting is the queue in front of a ticket purchase counter in a railway station on the Mumbai suburban railway [22]. Consider the passengers waiting in a queue to purchase a ticket from a human server at the Churchgate station on the Western line. A passenger traveling from Churchgate to Dadar (approximately 11 km) is impatient, updates the belief on the human server's rate of service based on the observed queue length, and balks in the case of a long queue at the ticket purchase counter as the cost of switching to another mode of transport is not high. On the other hand, a passenger traveling from Churchgate to Virar (approximately 86 km) is patient, does not make any strategic decision, and always joins the queue as the alternative modes of transport are expensive/tiresome.

McDonald's is another example of a queueing system with a human server offering counter service. Two types of customers arrive at the facility to order food. A patient customer is fond of hamburgers and will consume a hamburger with certainty in this visit. An impatient customer, on the contrary, is not particular on a McDonald's hamburger. Hence, an impatient customer updates the belief on the human server's service rate based on the observed queue length. Further, the impatient customer joins the queue if it is valuable enough to consume the hamburger, given the waiting cost involved; otherwise, the impatient customer balks. In a certain sense, patient and impatient customers are similar to loyal and disloyal customers, respectively.

As evident from the motivating examples, the patient customers are non-strategic because such customers possess infinite valuation for the service of interest. They always join the queue as they derive higher utility from joining than balking. Thus, non-strategic customers are equivalent to strategic customers with infinite service valuation. In this article, we assume that the patient customers are non-strategic and always join the queue. Further, it is pertinent to note that the finite waiting area capacity always does not ensure a finite queue length in practice at a Mumbai suburban railway station and a McDonald's outlet. Though the waiting area capacity is finite, queue length can sometimes explode, resulting in customers waiting outside the waiting area. While our primary focus is on analyzing the finite waiting area capacity model in this article, we also address the infinite waiting area capacity case in Sect. 7.

In this queueing environment, we develop a symmetric game amongst the impatient customers, where each impatient customer formulates the best response strategy in response to other impatient customers' strategies. Though patient customers are non-strategic, we assert that their presence impacts the queue length distribution and, therefore, the impatient customers' equilibrium strategies. Our results suggest that there always exists at least one pure threshold strategy equilibrium at which the impatient customers balk in the impatient customers' game in the presence of the patient customers. We then analyze the effect of the proportion of impatient customers and

the waiting area capacity on the likelihood ratio of a fast server to a slow server in an empty queue and the minimal pure threshold strategy equilibrium. As multiple pure threshold strategy equilibria can exist in some instances, we consider the minimal pure threshold strategy equilibrium, which is the minimum of the pure threshold strategy equilibria. We show that the likelihood ratio is monotonically decreasing in the proportion of impatient customers and monotonically increasing in the waiting area capacity. Further, we show that the minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers and non-decreasing in the waiting area capacity. We also observe that at least one pure threshold strategy equilibrium exists in an infinite capacity queue. We perform numerical analysis to analyze the effect of the proportion of impatient customers and the waiting area capacity on the expected throughput and the expected blocking rate.

In Sect. 2, we provide a review of the literature on queueing games and position our work in the context of the existing literature. We then describe the modeling environment in Sect. 3 and analyze the equilibrium structure in the impatient customers' game in Sect. 4. In the subsequent section (Sect. 5), we perform sensitivity analysis and study the effect of the proportion of impatient customers and the waiting area capacity on the likelihood ratio and the minimal pure threshold strategy equilibrium. We then perform a numerical analysis of the system performance in Sect. 6. We discuss the case of infinite waiting area capacity in Sect. 7. Finally, we provide concluding remarks and suggest promising future research directions in Sect. 8.

2 Related work

The notion of strategic behavior in observable queues dates back to Naor [23]. Hassin and Haviv [10] and Hassin [11] present a detailed review of the literature on strategic behavior in queues. We draw our work from two different dimensions of literature on queueing games—(1) service type, and (2) customer's strategic decision.

2.1 Service type

The extant literature on queueing games has considered various service types such as expert services where the service quality is not exactly known even after the service is provided [3], customer-intensive services where service value increases with service time [1], queues with catastrophes where a catastrophe forces the customers to abandon the system and makes the system inoperative for an exponential amount of time until the repair is done [2], and partial breakdowns where the rate of service comes down during a breakdown [19] among others. In recent times, there has been a surge in the literature on positive aspects of waiting. Debo et al. [4] developed a queueing game to signal service quality via queue lengths to uninformed customers in an environment consisting of both informed and uninformed customers. Meanwhile, Anand et al. [1] considered customer-intensive services such as medical, legal, and consultancy services where service time is positively correlated with service value, i.e., service value increases with service time.

Subsequently, Debo and Veeraraghavan [5] analyzed a queueing game amongst the customers in a customer-intensive environment when the service times are unknown, and the arriving customers update their beliefs on the service time based on observed queue length and then make queue-joining decisions. Recently, Liu and Shang [21] considered a queueing game amongst the customers when the service rates are unknown in a non-customer-intensive service environment where the service time is negatively correlated with service value. Our contribution to this strand of literature is similar to Liu and Shang [21] wherein the services are non-customer-intensive. However, the ensuing queueing game amongst the impatient customers in our case is not straightforward and perturbed by the presence of the patient customers. The proportion of patient customers influences the queue length distribution and, subsequently, the Bayesian updating in the utility functions of the strategic impatient customers. Therefore, it is of particular interest to understand the resultant equilibrium structure in this context.

2.2 Strategic customer decisions

Hassin [11] listed the various customer decisions in a queue such as balking, renegeing, arrival time, jockeying, retrials, restart. Queueing games with strategic balking customers where the arriving customers decide whether to join the queue or balk in the current period have been studied thoroughly in different contexts [6,8]. The other strategic customer decisions previously analyzed include arrival time, where the customers decide on the time to arrive at a queue [13,15], and retrials, where the customer comes back after some time if the server is busy [7,18] among others. There is also enough literature on loss systems wherein the arrivals are blocked in the presence of busy servers or finite waiting area capacity [11,24].

Our focus is on a non-customer-intensive service environment with unknown service rates consisting of impatient customers who make strategic balking decisions and patient customers who always join the queue and do not make any decision. As mentioned earlier, impatient customers have been studied extensively in the literature. With respect to patient customers, there have been attempts at modeling customers' patience in various environments different from ours. Iravani and Balcioglu [14] considered different classes of customers who exercise the renegeing option when their waiting time exceeds the patience limit. Liu and Cooper [20] then studied a single product pricing problem consisting of impatient and patient customers. While a patient customer waits up to some time periods for the product price to fall below his/her valuation to make a purchasing decision, an impatient customer decides to purchase immediately or leaves without buying. Our work is closely related to Hassin and Roet-Green [12] who study the effect of inspection cost (the cost of acquiring queue length information) on revenue and social welfare in a single server queue consisting of urgent and non-urgent customers. While urgent customers always join the queue, non-urgent customers join the queue according to their equilibrium strategy. We can infer that impatient and patient customers in our model are similar to non-urgent and urgent customers in Hassin and Roet-Green [12]. However, we analyze the strategic impatient customers'

equilibrium balking decisions in the non-strategic patient customers' presence in a non-customer-intensive service environment with unknown service rates.

2.3 Contribution to the queueing games literature

Our contribution to the literature on strategic queueing with unknown service rates is closely related, albeit different from Naor [23], Debo and Veeraraghavan [5] and Liu and Shang [21]. While impatient customers who make balking decisions have been studied in [5,21,23], we analyze the effect of patient customers who always join the queue (similar to urgent customers in Hassin and Roet-Green [12]) on the equilibrium strategies of impatient customers. In particular, we study the queueing game amongst the impatient customers when the service rates are unknown, similarly to Debo and Veeraraghavan [5] and Liu and Shang [21]. Further, we consider non-customer-intensive services similar to Liu and Shang [21] and unlike Debo and Veeraraghavan [5]. The comparison between our model, Naor [23], Debo and Veeraraghavan [5] and Liu and Shang [21] is presented in Table 1.

3 Model and notation

3.1 Notation

Λ	Arrival rate
$\bar{\mu}$	Fast service rate
$\underline{\mu}$	Slow service rate
p	Probability (prior) that the expected service rate is fast
q	Proportion of impatient customers
V	Service value
c	Waiting cost per unit time
v_n	Impatient customer's posterior belief that the service rate is fast when the queue length is n
ψ_n	Impatient customer's joining probability when the queue length is n
n_b	Highest balking threshold for an impatient customer
\bar{N}	Capacity of the waiting area
v	Impatient customer's posterior belief vector
ψ	Impatient customer's joining strategy vector
$U(n, v)$	Impatient customer's utility as a function of queue length and posterior belief
$U(n, \psi, q)$	Impatient customer's utility as a function of queue length, impatient customer's joining probability and proportion of impatient customers
$\pi_{n,\mu,\psi,q}$	Stationary probability of n customers in the system as a function of service rate, impatient customer's joining probability and proportion of impatient customers
$b(k)$	Best response of a focal impatient customer when the other customers join according to a pure threshold strategy k

Table 1 Our model versus Naor [23] versus Debo and Veeraraghavan [5] versus Liu and Shang [21]

	[23]	[5]	[21]	Our model
Service time - Service value correlation	Negative	Positive	Negative	Negative
Game amongst the customers	×	✓	✓	✓
Strategic impatient customer	✓	✓	✓	✓
Non-strategic patient customer	×	×	×	✓

3.2 Model description

Nature first chooses the mean service rate $\mu \in \{\underline{\mu}, \bar{\mu}\}$, where $\underline{\mu} < \bar{\mu}$. The expected service rate is the fast $\bar{\mu}$ with probability p and the slow $\underline{\mu}$ with probability $1 - p$, which is assumed to be common knowledge. Customers arrive at the service facility according to a Poisson process with parameter Λ . The arriving customers are of two types. A proportion q of customers is impatient and make balk or join decisions, and the remaining $1 - q$ proportion of customers is patient and always joins the queue. The service times follow an exponential distribution. An impatient customer observes a queue of length n and updates the belief on the service rate. The posterior belief on service rate is described as follows: the impatient customer believes with probability v_n that the service rate is fast and probability $1 - v_n$ that the service rate is slow. The impatient customer obtains a service value V and incurs a waiting cost of $\frac{n + 1}{\mu}c$, where n is the observed queue length on arrival (including the customer in service), and c is the waiting cost per unit time. When the queue length is n^* , we assume that the impatient customer obtains positive utility if the service rate is $\bar{\mu}$, $V - \frac{(n^* + 1)c}{\bar{\mu}} > 0$, and obtains negative utility if the service rate is $\underline{\mu}$, $V - \frac{(n^* + 1)c}{\underline{\mu}} < 0$.

Assumption 1 There exists $n^* \geq 0$ such that $V\underline{\mu} < (n^* + 1)c < V\bar{\mu}$.

Therefore, the impatient customer’s utility function is defined as

$$U(n, v) = v_n \left(V - \frac{n + 1}{\bar{\mu}}c \right) + (1 - v_n) \left(V - \frac{n + 1}{\underline{\mu}}c \right). \tag{1}$$

We normalize the utility from balking to 0. Let $n_b = \left\lfloor \frac{V\bar{\mu}}{c} \right\rfloor$, where n_b is the highest balking threshold for the impatient customer. To reflect the waiting area capacity in reality, we define $\bar{N} = w$, where $w \in \mathbb{N}$. However, we will show that the existence of equilibrium is guaranteed even when the waiting area capacity is infinite in Sect. 7. At \bar{N} , the customers are blocked from entering the system. To avoid cases where the impatient customer is blocked from the system, we make the following assumption.

Assumption 2 We assume that $n_b < \bar{N}$.

n_b is a result of impatient customer’s strategic decision making and can be computed by the service provider. The service provider can thus set the capacity greater than n_b . There is abundant literature on queueing games with impatient customers who make balk or join decisions [11]. In such a setting, the notion of blocking does not arise because n_b is the highest balking threshold and $n_b < \bar{N}$. On the other hand, patient customers always join the queue in a queueing system consisting of impatient and patient customers. Thus, queue lengths beyond n_b are attained with positive probability in the presence of patient customers. Subsequently, patient customers are blocked from entering the system at \bar{N} .

While the state space of observed queue lengths is $n \in \{0, 1, 2, \dots, \bar{N}\}$, the equilibrium analysis in the impatient customer’s game is restricted to $n \in \{0, 1, 2, \dots, n_b\}$. $\psi_n \in [0, 1]$ denotes the impatient customer’s joining probability after observing a queue of length n for all $n \in \{0, 1, 2, \dots, n_b\}$. We use the vectors $\psi = (\psi_0, \psi_1, \dots, \psi_{n_b})$ and $v = (v_0, v_1, \dots, v_{n_b})$ to represent the impatient customer’s joining strategy and updated belief, respectively. We define $\pi_{n,\mu,\psi,q}$ as the stationary probability of exactly n customers in the system when nature chooses service rate μ , proportion q of the customers are impatient and impatient customers play a strategy ψ for all $n \in \{0, 1, 2, \dots, n_b\}$.

Lemma 1 *The stationary probability of n customers in the system $\pi_{n,\mu,\psi,q}$ is given by*

$$\pi_{n,\mu,\psi,q} = \begin{cases} \pi_{0,\mu,\psi,q} \prod_{j=0}^{n-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{1, 2, \dots, n_b\}, \\ \pi_{0,\mu,\psi,q} \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \prod_{j=0}^{n_b-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{n_b + 1, n_b + 2, \dots, \bar{N}\}, \end{cases} \tag{2}$$

where $\pi_{0,\mu,\psi,q}$ is the stationary probability of an empty system, which in turn is equal to

$$\pi_{0,\mu,\psi,q} = \left[1 + \sum_{n=1}^{n_b} \prod_{j=0}^{n-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right) + \prod_{j=0}^{n_b-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right) \sum_{n=n_b+1}^{\bar{N}} \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \right]^{-1}. \tag{3}$$

The proofs of all lemmas are available in the appendix. We observe that the stationary probability of n customers depends on the impatient customer’s balking strategy and the proportion of impatient and patient customers. We use the PASTA property (Poisson Arrivals See Time Averages), which states that a randomly arriving customer in the limit observes a queue of length n with stationary probability $\pi_{n,\mu,\psi,q}$ [25]. According to Bayes’ rule, the posterior probability that the service rate is fast is given by

$$v_n = \frac{p\pi_{n,\bar{\mu},\psi,q}}{p\pi_{n,\bar{\mu},\psi,q} + (1-p)\pi_{n,\underline{\mu},\psi,q}}. \quad (4)$$

4 Equilibrium analysis in the impatient customers' game

We determine the equilibrium strategies ψ^* and updated equilibrium beliefs v^* of the impatient customer. In the following definition, we specify the equilibrium conditions in the customer's game.

Definition 1 The impatient customers' strategies ψ^* and updated beliefs v^* form an equilibrium under the following conditions:

(i) The impatient customers are rational. For all $n \in \{0, 1, 2, \dots, n_b\}$,

$$\psi_n^* \in \arg \max_{\tilde{\psi} \in [0,1]} \tilde{\psi} U(n, \psi^*, q). \quad (5)$$

(ii) The beliefs of the impatient customers are consistent. Under the customer's strategy ψ^* , the belief v_n^* satisfies Bayes' rule for all queue lengths $n \in \{0, 1, 2, \dots, n_b\}$.

We can express the utility function $U(n, v)$ as a function of strategies ψ , and proportion of impatient customers q , by substituting v_n in $U(n, v)$:

$$U(n, \psi, q) = \frac{p\pi_{n,\bar{\mu},\psi,q} \left(V - \frac{n+1}{\bar{\mu}} c \right) + (1-p)\pi_{n,\underline{\mu},\psi,q} \left(V - \frac{n+1}{\underline{\mu}} c \right)}{p\pi_{n,\bar{\mu},\psi,q} + (1-p)\pi_{n,\underline{\mu},\psi,q}}. \quad (6)$$

The equilibrium conditions for an impatient customer expressed in terms of ψ^* are given by

$$\psi_n^* = 1 \iff U(n, \psi^*, q) > 0, \quad (7)$$

$$\psi_n^* = 0 \iff U(n, \psi^*, q) < 0, \quad (8)$$

$$\psi_n^* \in (0, 1) \iff U(n, \psi^*, q) = 0. \quad (9)$$

As the action space $[0, 1]^{n_b}$ is compact and the best response function graph is closed, we conclude from the Kakutani fixed-point theorem that at least one equilibrium in mixed strategies exists [16]. Our equilibrium analysis is only applicable on queue lengths $n \in \{0, 1, \dots, n_b\}$.

The resulting equilibrium structure can be of threshold or non-threshold type. A threshold strategy is described by the threshold $k = n + z$, where $n \in \mathbb{N}$ and $z \in [0, 1)$. The customers join when the queue length is $0 \leq i \leq n - 1$; join with probability z and do not join with probability $1 - z$ when the queue length is n ; do not join when $i > n$. If $z = 0$ (k is an integer), it is a pure strategy; otherwise, it is mixed [10]. On the other hand, a non-threshold strategy describes the non-monotone joining behavior of customers and is not solely defined by the threshold. Some examples of non-threshold strategies include a sputtering strategy [5] and strategy with a hole [4].

In Debo and Veeraraghavan [5], a sputtering strategy is defined by two queue lengths \underline{k} and \bar{k} where the customers join when the queue length is $0 \leq i \leq \underline{k} - 1$; join with probability $\alpha_{\underline{k}}$ and do not join with probability $1 - \alpha_{\underline{k}}$ at queue length \underline{k} ; join again at queue lengths $\underline{k} + 1 \leq i < \bar{k}$; balk at queue length $\bar{k} > \underline{k} + 1$. Debo et al. [4] consider a non-threshold strategy with a hole for an uninformed customer in a population mix of informed and uninformed customers. At the hole \hat{n} , the uninformed customer does not join the queue; however, the uninformed customer follows the informed customer’s strategy at all queue lengths except \hat{n} . The informed customer joins at queue lengths $i \geq \hat{n}$ for a high-quality firm. The uninformed customer’s strategy is of non-threshold type as the uninformed customer balks at \hat{n} and joins at queue lengths greater than \hat{n} .

We perform the equilibrium analysis by adopting the framework of Debo and Veeraraghavan [5] to our context where the services are non-customer-intensive, and customers are of two types. Our focus is on threshold equilibrium structures for the queueing game where a pure or mixed strategy equilibrium exists. We first consider pure threshold strategies where the arriving impatient customers join at the first $k - 1$ queue lengths and balk at k . At queue lengths $\{k, k + 1, \dots, \bar{N} - 1\}$, only the patient customers join the queue. A threshold strategy k is an equilibrium if it satisfies the following conditions: $U(n, k, q) \geq 0$ for all $0 \leq n \leq k - 1$

and $U(n, k, q) \leq 0$ for all $k \leq n \leq n_b$. Let $\omega(n) = \left(\frac{\bar{\mu}}{\mu}\right)^n \frac{\mu}{V - \frac{n+1}{\bar{\mu}}c} - V$

$$\theta_q(k, \bar{N}) = \frac{1 + \sum_{j=1}^k \left(\frac{\Lambda}{\mu}\right)^j + \sum_{j=k+1}^{\bar{N}} (1 - q)^{j-k} \left(\frac{\Lambda}{\mu}\right)^j}{1 + \sum_{j=1}^k \left(\frac{\Lambda}{\bar{\mu}}\right)^j + \sum_{j=k+1}^{\bar{N}} (1 - q)^{j-k} \left(\frac{\Lambda}{\bar{\mu}}\right)^j}.$$

Lemma 2 *The pure threshold strategy k is in equilibrium if and only if, $\omega(n) \leq \frac{p}{1-p} \theta_q(k, \bar{N})$ for $n \in \{0, 1, \dots, k - 1\}$ and $\omega(n) \geq \frac{p}{1-p} \theta_q(k, \bar{N})$ for $n \in \{k, k + 1, \dots, n_b\}$.*

We now consider $\omega(n)$ and $\theta_q(k, \bar{N})$ as a function over the real numbers, and replace integer n by real \tilde{n} and integer k by real \tilde{k} . Further, let $\underline{n} = \frac{V\mu}{c} - 1$ and $\bar{n} = \frac{V\bar{\mu}}{c} - 1$. We also notice that $\omega(\tilde{n})$ is non-negative for $\tilde{n} \in [\underline{n}, \bar{n}]$ and $\omega(\underline{n}) = 0$ and $\lim_{\tilde{n} \rightarrow \bar{n}} \omega(\tilde{n}) = +\infty$. In addition, $[\underline{n}, \bar{n}]$ is non-empty and the domain of $\omega(\tilde{n})$ is $[\underline{n}, \bar{n}]$. As $\theta_q(\tilde{k}, \bar{N})$ is non-negative for all $\tilde{k} \geq 0$, the domain of $\theta_q(\tilde{k}, \bar{N})$ is \mathbb{R}^+ . We then define $N(\tilde{k}) = \min \left\{ \tilde{n} \in [\underline{n}, \bar{n}] : \omega(\tilde{n}) \geq \xi \theta_q(\tilde{k}, \bar{N}) \right\}$ where $\xi = \frac{p}{1-p}$. The domain of $N(k)$ is not restricted to $[0, k]$. Then, $b(k) = \max\{0, \lceil N(k) \rceil\}$ is the best response of a focal customer when the other customers join based on a pure threshold strategy k . When the best response of a focal customer is equal to the strategy of other customers, i.e., $b(k) = k$ or $\omega(k - 1) < \xi \theta_q(k, \bar{N}) < \omega(k)$, then the focal customer’s strategy $b(k)$ is an equilibrium strategy.

Remark 1 At $q = 0$, $\theta_q(\tilde{k}, \bar{N})$ is independent of \tilde{k} and all customers are patient (non-strategic) and join the queue until \bar{N} at which they are blocked.

Remark 2 At $q = 1$, the focal impatient customer’s best response depends on other

customers’ threshold balking strategies and $\theta_q(\tilde{k}, \bar{N})$ reduces to $\frac{1 + \sum_{j=1}^{\tilde{k}} \left(\frac{\Lambda}{\underline{\mu}}\right)^j}{1 + \sum_{j=1}^{\tilde{k}} \left(\frac{\Lambda}{\bar{\mu}}\right)^j}$,

which is the inverse of $\Phi(\tilde{k})$ in Debo and Veeraraghavan [5]. It follows that $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \tilde{k} .

In our case, $\theta_q(k, \bar{N}) = \frac{\pi_{0, \bar{\mu}, k, q}}{\pi_{0, \underline{\mu}, k, q}}$ represents the likelihood ratio of a fast server to a slow server for an empty queue, which is the opposite of the likelihood ratio defined in Debo and Veeraraghavan [5]; however, $1 - q$ proportion of the customers are patient in our setting. On the other hand, we consider the likelihood ratio for a queue of length n as $\frac{\pi_{n, \bar{\mu}, k, q}}{\pi_{n, \underline{\mu}, k, q}} = \left(\frac{\underline{\mu}}{\bar{\mu}}\right)^n \theta_q(k, \bar{N})$. This holds under the assumption that customers join at queue lengths between 0 and $n - 1$. An arriving impatient customer is ambivalent between joining and not when $\left(\frac{n+1}{\underline{\mu}}c - V\right) / \left(V - \frac{n+1}{\bar{\mu}}c\right) = \frac{p}{1-p} \left(\frac{\underline{\mu}}{\bar{\mu}}\right)^n \omega(n)$. Similar to Debo and Veeraraghavan [5], we can interpret $\omega(n)$ as the required likelihood ratio for an empty queue that makes the arriving customer ambivalent at queue length n .

Lemma 3 (i) If $q > 0$, $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \tilde{k} for all $\tilde{k} \geq 0$.
 (ii) $\omega(\tilde{n})$ is monotonically increasing in \tilde{n} over $[\underline{n}, \bar{n}]$.

We show that the queue joining strategies of impatient customers are monotone in the queue length when there exists a negative correlation between service time and service value for all $q > 0$. We observe the follow-the-crowd (FTC) behavior and then show that at least one pure threshold strategy equilibrium is guaranteed. Recently, Liu and Shang [21] proved the existence of at least one threshold equilibrium in a setting consisting of only impatient customers. However, we show that at least one pure threshold equilibrium strategy exists even in the presence of patient customers (Proposition 1). This result is also different from Naor’s model with negative externalities which exhibits neither the follow-the-crowd (FTC) nor the avoid-the-crowd (ATC) behavior where a customer’s decision is independent of the other customers’ decisions [11].

Proposition 1 If $q > 0$, there exists at least one pure threshold strategy equilibrium in the impatient customer’s game.

Proof Let $k_1 > k_2$. From Lemma 3, we observe that $\theta_q(k_1) > \theta_q(k_2)$ for all $q > 0$. Also, $\omega(n)$ is monotonically increasing in n . By the definition of $b(k)$, it is clear that $\theta_q(k_1) > \theta_q(k_2) \implies N(k_1) \geq N(k_2)$. We then find that $b(k_1) \geq b(k_2)$ when

$k_1 > k_2$ which demonstrates the FTC behavior. The higher the threshold adopted by other impatient customers, the higher is the focal impatient customer’s threshold to balk. We know that multiple fixed points exist and therefore, multiple symmetric equilibria of the threshold type in pure or mixed strategy space are possible in the presence of FTC behavior [10].

Further, if $\omega(\lceil n \rceil) \geq \xi\theta_q(\lceil n \rceil)$, $\lceil n \rceil$ is a pure threshold strategy equilibrium. Instead, if $\omega(\lceil n \rceil) < \xi\theta_q(\lceil n \rceil)$, we will show that there always exists a pure threshold strategy equilibrium at $k > \lceil n \rceil$. We know that $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \tilde{k} and $\omega(\tilde{n})$ is monotonically increasing in \tilde{n} (Lemma 3). However, $\lim_{\tilde{n} \rightarrow \bar{n}} \omega(\tilde{n}) = +\infty$. Therefore, there always exists a \tilde{k}' at which the functions $\xi\theta_q(\tilde{k}', \bar{N})$ and $\omega(\tilde{k}')$ intersect. Let $\lceil \tilde{k}' \rceil = k$. We then know that $\omega(k - 1) < \xi\theta_q(k - 1, \bar{N})$ and $\xi\theta_q(k, \bar{N}) < \omega(k)$. Moreover, as both $\theta_q(\tilde{k}, \bar{N})$ and $\omega(\tilde{k})$ are monotonically increasing in \tilde{k} (Lemma 3), $\omega(k - 1) < \xi\theta_q(k, \bar{N})$. Hence, $\omega(k - 1) < \xi\theta_q(k, \bar{N}) < \omega(k)$ or $b(k) = k$ and at least one pure threshold strategy equilibrium exists. \square

In the equilibrium analysis of priority queues, [9] illustrates the existence of a mixed threshold equilibrium strategy between two consecutive pure threshold equilibrium strategies. We also observe a similar phenomenon in our case when consecutive pure threshold strategy equilibria exist. Due to the FTC behavior in our queueing environment, multiple threshold equilibria are possible. When there are two consecutive pure threshold strategy equilibria (i.e., $\xi\theta_q(k - 1, \bar{N}) < \omega(k - 1) < \xi\theta_q(k, \bar{N}) < \omega(k)$), we will show that there exist infinitely many mixed threshold strategy equilibria. The impatient customers join at queue length $k - 1$ with probability ψ_{k-1} and balk at k . The queue-joining probability ψ_{k-1} satisfies the condition $\omega(k - 2) < \frac{p}{1 - p} \hat{\theta}_q(k - 1, \psi_{k-1}, k, \bar{N}) < \omega(k)$, where

$$\hat{\theta}_q(k - 1, \psi_{k-1}, k, \bar{N}) = \frac{\sum_{j=0}^{k-1} \left(\frac{\Lambda}{\mu}\right)^j + (\psi_{k-1}q + 1 - q) \left[\left(\frac{\Lambda}{\mu}\right)^k + \sum_{j=k+1}^{\bar{N}} (1 - q)^{j-k} \left(\frac{\Lambda}{\mu}\right)^j \right]}{\sum_{j=0}^{k-1} \left(\frac{\Lambda}{\mu}\right)^j + (\psi_{k-1}q + 1 - q) \left[\left(\frac{\Lambda}{\mu}\right)^k + \sum_{j=k+1}^{\bar{N}} (1 - q)^{j-k} \left(\frac{\Lambda}{\mu}\right)^j \right]}$$

Proposition 2 *If there exist pure threshold strategy equilibria at consecutive queue lengths $k - 1$ and k , there are infinitely many mixed threshold strategy equilibria with randomization at $k - 1$.*

Proof The definition of pure threshold strategy equilibria at queue lengths $k - 1$ and k implies that the conditions $\omega(k - 2) < \xi\theta_q(k - 1, \bar{N}) < \omega(k - 1)$ and $\omega(k - 1) < \xi\theta_q(k, \bar{N}) < \omega(k)$ are satisfied, respectively. Further, we can infer that $\xi\theta_q(k - 1, \bar{N}) < \omega(k - 1) < \xi\theta_q(k, \bar{N})$. Using the definition of $\hat{\theta}_q(k - 1, \psi_{k-1}, k, \bar{N})$, it follows that $\xi\hat{\theta}_q(k - 1, 0, k, \bar{N}) < \omega(k - 1) < \xi\hat{\theta}_q(k - 1, 1, k, \bar{N})$. As $\hat{\theta}_q(k - 1, \psi_{k-1}, k, \bar{N})$ is continuous in ψ_{k-1} , there always exists infinitely many $\psi_{k-1} \in [0, 1]$ at which

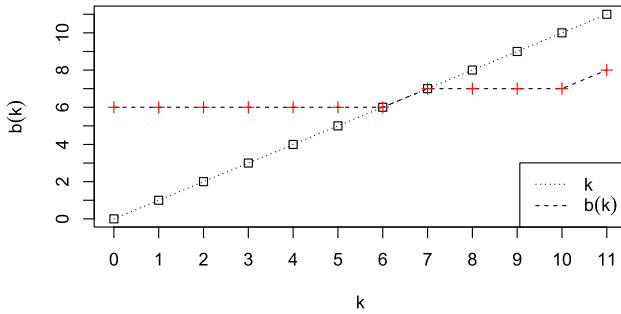


Fig. 1 Two consecutive pure threshold strategy equilibria at $b(k) = k = 6$ and $b(k) = k = 7$, respectively

Table 2 Focal impatient customer’s best response $b(k)$ to the threshold strategy k

k	$\xi\theta_q(k, \bar{N})$	$\omega(k)$	$b(k)$
0	2.727	–	6
1	3.344	–	6
2	4.490	–	6
3	6.333	–	6
4	9.204	–	6
5	13.631	0	6
6	20.423	25.6	6
7	30.801	128	7
8	46.569	512	7
9	70.334	2048	7
10	105.740	10240	7
11	157.608	∞	8

$\omega(k - 2) < \hat{\theta}_q(k - 1, \psi_{k-1}, k, \bar{N}) < \omega(k)$. Also, with monotonically increasing $\omega(n)$ (Lemma 3), we know that $\omega(n) < \omega(k - 1)$ for all $0 \leq n < k - 1$. This justifies the existence of infinitely many mixed threshold strategy equilibria with queue-joining probability ψ_{k-1} at $k - 1$ and balking at k . \square

We provide the intuition behind the existence of infinitely many equilibria with any randomization ψ_{k-1} at $k - 1$. A pure threshold strategy equilibrium at k implies that the impatient customers join at $k - 1$ with probability 1. Similarly, a pure strategy equilibrium at $k - 1$ implies that the impatient customers join at $k - 1$ with probability 0. As both joining and balking at $k - 1$ satisfy equilibrium conditions, any joining probability at $k - 1$ will satisfy the equilibrium conditions. Hence, there exist infinitely many mixed threshold strategy equilibria.

To illustrate the existence of infinitely many mixed threshold strategy equilibria between any two consecutive pure threshold strategy equilibria, we provide the following example. Let $p = 0.05$, $\Lambda = 15$, $\bar{\mu} = 12$, $\underline{\mu} = 6$, $q = 0.5$, $\bar{N} = 15$, $V = 2$, $c = 2$. In Fig. 1, we illustrate the existence of consecutive pure threshold strategy equilibria at $b(k) = k = 6$ and $b(k) = k = 7$, respectively. We observe from Table 2 that $\omega(5) < \xi\theta_{0.5}(6, 15) < \omega(6) < \xi\theta_{0.5}(7, 15) < \omega(7) \implies \omega(5) <$

$\xi \hat{\theta}_{0,5}(6, 0, 7, 15) < \omega(6) < \xi \hat{\theta}_{0,5}(6, 1, 7, 15) < \omega(7)$. There exist infinitely many mixed threshold strategy equilibria at which $\omega(5) < \hat{\theta}_{0,5}(6, \psi_1, 7, 15) < \omega(7)$.

5 Sensitivity analysis

We perform sensitivity analysis by studying the impact of the proportion of impatient customers and the waiting area capacity on threshold equilibrium strategies. As multiple pure threshold strategy equilibria can possibly exist in certain cases, we particularly consider the effect of the proportion of impatient customers and the waiting area capacity on the minimal pure threshold strategy equilibrium. We first analyze the effect of the proportion of impatient customers on impatient customers’ equilibrium strategies. In this regard, we aim to understand how $\theta_q(k, \bar{N})$ varies with the proportion of impatient customers (Lemma 4).

Lemma 4 $\theta_q(\tilde{k}, \bar{N})$ is monotonically decreasing in q for all $\tilde{k} \geq 0$.

Proposition 3 The minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers.

Proof From Lemma 4, we know that $\theta_{q_1}(k, \bar{N}) < \theta_{q_2}(k, \bar{N})$ when $q_1 > q_2$. As $\omega(n)$ is independent of q and monotonic, it follows that $N_{q_1}(k) \leq N_{q_2}(k) \implies b_{q_1}(k) \leq b_{q_2}(k)$ for a given k . As a pure threshold strategy equilibrium is guaranteed to exist (Proposition 1), let the equilibrium for q_2 occur at k , i.e., $b_{q_2}(k) = k$. If multiple equilibria exist for q_2 , consider the smallest pure threshold strategy equilibrium. Now, as $b_{q_1}(k) \leq b_{q_2}(k)$, we can look at two cases:

Case 1. If $b_{q_1}(k) = b_{q_2}(k)$, then the equilibrium for q_1 also occurs at k .

Case 2. If $b_{q_1}(k) < b_{q_2}(k)$, the equilibrium for q_1 does not occur at k . From the FTC behavior, we know that $b_{q_1}(k) \geq b_{q_1}(k - 1)$. Now, if $b_{q_1}(k - 1) = k - 1$, the pure threshold strategy equilibrium for q_1 occurs at $k - 1$. Otherwise, there is no pure threshold strategy equilibrium at $k - 1$ for q_1 . Then, we can consider $k - 2, k - 3, \dots, \lfloor \underline{n} \rfloor$. In general, $b_{q_1}(k - \varphi)$ is bounded below by $\lfloor \underline{n} \rfloor$ for all $\varphi \in \{0, 1, 2, \dots, k\}$. As $b_{q_1}(k - \varphi + 1) \geq b_{q_1}(k - \varphi)$, we find that $b_{q_1}(k - \varphi) = \lfloor \underline{n} \rfloor$ if $b_{q_1}(k - \varphi + 1) = \lfloor \underline{n} \rfloor$. Therefore, if there exists no pure threshold strategy equilibrium from $k - 1$ (i.e., $b_{q_1}(k - 1) \neq k - 1$) to $\lfloor \underline{n} \rfloor + 1$ (i.e., $b_{q_1}(\lfloor \underline{n} \rfloor + 1) \neq \lfloor \underline{n} \rfloor + 1$), there will always exist a pure threshold strategy equilibrium for q_1 at $\lfloor \underline{n} \rfloor$ (i.e., $b_{q_1}(\lfloor \underline{n} \rfloor) = \lfloor \underline{n} \rfloor$). \square

At this stage, it is essential to recall the physical interpretation of $\theta_q(k, \bar{N})$, which is the likelihood ratio of a fast server to a slow server in an empty queue. We also know that an impatient customer prefers a fast server over a slow server when the queue length is 0. As the likelihood ratio is monotonically decreasing in the proportion of impatient customers, it follows that the minimal symmetric pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers. From Proposition 3, we can also infer that the probability of attaining higher queue lengths is low when the proportion of impatient customers is high. Conversely, the probability of reaching higher queue lengths is high when the proportion of patient customers is high. The reason behind this behavior is that the minimal pure threshold strategy equilibrium

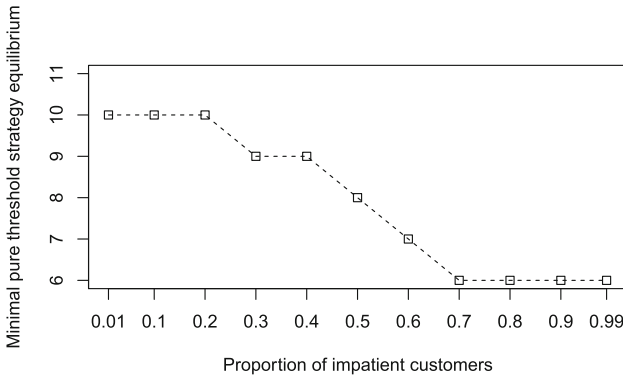


Fig. 2 Minimal pure threshold strategy equilibrium versus proportion of impatient customers

and the joining rate are non-decreasing in the proportion of patient customers. We provide an example to show that the minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers where $p = 0.25$, $\Lambda = 15$, $\bar{\mu} = 12$, $\underline{\mu} = 6$, $\bar{N} = 15$, $V = 2$, $c = 2$ (refer to Fig. 2).

We next study how the pure threshold strategy equilibrium varies with the waiting area capacity.

Lemma 5 $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \bar{N} for all $\tilde{k} \geq 0$.

Proposition 4 The minimal pure threshold strategy equilibrium is non-decreasing in the capacity of the waiting area.

Proof This proof is similar to the proof of Proposition 3 and is provided in the appendix. □

As the likelihood ratio is monotonically increasing in the waiting area capacity, it follows that the minimal symmetric pure threshold strategy equilibrium is non-decreasing in the waiting area capacity. Proposition 4 is significant because the equilibrium in the impatient customers’ game occurs at potentially higher queue lengths when the waiting area capacity is high. We provide an example where $p = 0.25$, $\Lambda = 15$, $\bar{\mu} = 12$, $\underline{\mu} = 6$, $q = 0.5$, $V = 2$, $c = 2$ (refer to Figure 3) to illustrate that the minimal pure threshold strategy equilibrium is non-decreasing in the waiting area capacity.

6 Numerical analysis of the system performance

We perform numerical analysis to analyze how the proportion of impatient customers and the waiting area capacity influence the system performance. To understand the system performance, we first define throughput for a server of type μ as $TP_{\mu, \psi, q, \bar{N}} = \mu(1 - \pi_{0, \mu, \psi, q})$, where $(1 - \pi_{0, \mu, \psi, q})$ denotes the probability that the server is busy

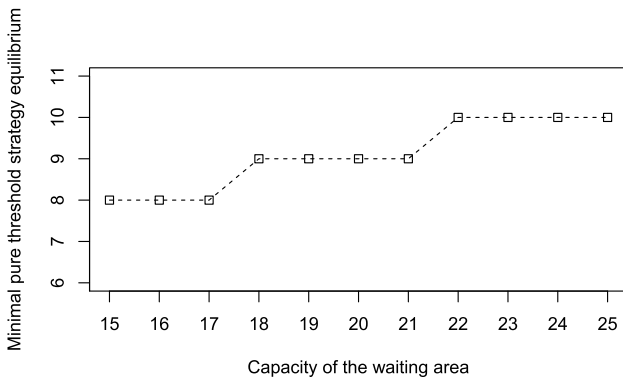


Fig. 3 Minimal pure threshold strategy equilibrium versus capacity of the waiting area

[4,5]. As a pure threshold strategy equilibrium k always exists, throughput becomes

$$TP_{\mu,k,q,\bar{N}} = \mu \left(1 - \frac{1}{\sum_{j=0}^k \left(\frac{\Lambda}{\mu}\right)^j + \left(\frac{\Lambda}{\mu}\right)^k \sum_{j=k+1}^{\bar{N}} \left(\frac{(1-q)\Lambda}{\mu}\right)^{j-k}} \right). \tag{10}$$

As the server can be fast or slow, we consider the expected throughput $E[TP] = pTP_{\bar{\mu},k,q,\bar{N}} + (1-p)TP_{\mu,k,q,\bar{N}}$ similar to Debo and Veeraraghavan [5].

As the waiting area capacity is finite, customers are blocked from entering the system at \bar{N} [24]. From Assumption 2, it is apparent that only the patient customers are blocked from entering the system at \bar{N} in our case. The blocking probability for a server of type μ is then equal to

$$\pi_{\bar{N},\mu,k,q} = \frac{\left(\frac{\Lambda}{\mu}\right)^k \left(\frac{(1-q)\Lambda}{\mu}\right)^{\bar{N}-k}}{\sum_{j=0}^k \left(\frac{\Lambda}{\mu}\right)^j + \left(\frac{\Lambda}{\mu}\right)^k \sum_{j=k+1}^{\bar{N}} \left(\frac{(1-q)\Lambda}{\mu}\right)^{j-k}}. \tag{11}$$

As only the patient customers are blocked from entering the system, we define the blocking rate as $BR_{\mu,k,q,\bar{N}} = \Lambda(1-q)\pi_{\bar{N},\mu,k,q}$. We then consider the expected blocking rate $E[BR] = pBR_{\bar{\mu},k,q,\bar{N}} + (1-p)BR_{\mu,k,q,\bar{N}}$.

After setting $\bar{\mu} = 12$, $\mu = 6$, $V = 2$, $c = 2$, we perform numerical analysis by considering different values of $p = \{0.1, 0.9\}$, $\Lambda = \{3, 9, 15\}$, $q = \{0.1, 0.5, 0.9\}$ and $\bar{N} = \{15, 20\}$. In particular, we analyze how the expected throughput varies with the proportion of impatient customers and the waiting area capacity. We also study how the expected blocking rate varies with the proportion of impatient customers and the waiting area capacity. When multiple pure threshold strategy equilibria exist, we consider the minimal pure threshold strategy equilibrium. The results are presented in Tables 3 and 4.

Table 3 Variation of the system performance with Λ , \bar{N} and q at $p = 0.1$

Λ	\bar{N}	q	Expected throughput	Expected blocking rate
3	15	0.1	2.9961	1.43×10^{-5}
3	15	0.5	2.9858	4.04×10^{-8}
3	15	0.9	2.9798	4.15×10^{-15}
3	20	0.1	2.9961	2.65×10^{-7}
3	20	0.5	2.9858	3.95×10^{-11}
3	20	0.9	2.9798	1.30×10^{-21}
9	15	0.1	6.2778	1.9257
9	15	0.5	6.1795	0.0543
9	15	0.9	6.0972	1.04×10^{-8}
9	20	0.1	6.2887	1.8999
9	20	0.5	6.1820	0.0125
9	20	0.9	6.0972	7.88×10^{-13}
15	15	0.1	6.5882	6.9535
15	15	0.5	6.5561	1.5231
15	15	0.9	6.5121	2.58×10^{-6}
15	20	0.1	6.5952	6.9305
15	20	0.5	6.5663	1.4153
15	20	0.9	6.5121	2.52×10^{-9}

From the numerical analysis, we find that the expected throughput is monotonically non-increasing in the proportion of impatient customers. This follows from Proposition 3, which states that the minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers. Further, an increase in the proportion of impatient customers leads to a corresponding decrease in the proportion of patient customers who always join the queue. Since the joining rate is monotonically non-increasing in the proportion of impatient customers, the expected throughput is monotonically non-increasing in the proportion of impatient customers. On the other hand, the expected throughput is monotonically non-decreasing in the waiting area capacity since it follows from Proposition 4 that the minimal pure threshold strategy equilibrium is non-decreasing in the waiting area capacity.

We then find that the expected blocking rate is monotonically non-increasing in the proportion of impatient customers. This behavior follows from Proposition 3. Further, it is also due to the decrease in the proportion of patient customers who always join the queue. While the expected blocking rate is monotonically decreasing in the waiting area capacity in most cases, this is not always guaranteed. We provide an example to show that the expected blocking rate is not always monotonically decreasing in the waiting area capacity. Let $p = 0.05$, $\Lambda = 22$, $\bar{\mu} = 20$, $\mu = 2$, $q = 0.9$, $V = 10$, $c = 22$. At $\bar{N} = 14$, multiple pure threshold strategy equilibria occur at 1 and 2. Then, at $\bar{N} = 15$, multiple pure threshold strategy equilibria occur at 1, 2 and 3. Assume that the equilibrium occurs at $b(k) = k = 1$ when $\bar{N} = 14$ and at $b(k) = k = 3$ when

Table 4 Variation of the system performance with Λ , \bar{N} and q at $p = 0.9$

Λ	\bar{N}	q	Expected throughput	Expected blocking rate
3	15	0.1	2.9995	1.60×10^{-6}
3	15	0.5	2.9981	4.50×10^{-9}
3	15	0.9	2.9973	4.61×10^{-16}
3	20	0.1	2.9995	2.95×10^{-8}
3	20	0.5	2.9981	4.39×10^{-12}
3	20	0.9	2.9973	1.44×10^{-22}
9	15	0.1	8.6413	0.2317
9	15	0.5	8.5232	0.0110
9	15	0.9	8.4796	4.99×10^{-8}
9	20	0.1	8.6627	0.2145
9	20	0.5	8.5234	0.0024
9	20	0.9	8.4796	3.77×10^{-12}
15	15	0.1	11.3079	2.6843
15	15	0.5	11.2471	0.3761
15	15	0.9	11.0852	1.94×10^{-5}
15	20	0.1	11.3571	2.3711
15	20	0.5	11.2527	0.1803
15	20	0.9	11.0852	1.79×10^{-8}

$\bar{N} = 15$. We then observe that the expected blocking rate is equal to 0.2571 at $\bar{N} = 14$ and 0.2664 at $\bar{N} = 15$.

Comparing Tables 3 and 4, we find that the expected throughput is monotonically increasing in p . As the probability of the server being fast increases, it is intuitive that the expected throughput will increase. The expected blocking rate is monotonically decreasing in p because the customers are served fast with higher probability. It is also intuitive to observe that the expected throughput and the expected blocking rate are monotonically increasing in the arrival rate.

7 The case of infinite waiting area capacity

We now relax the assumption on finite waiting area capacity \bar{N} , and consider the case where the waiting area is of infinite capacity. We use the earlier notation with ∞ in the superscript to reflect the infinite waiting area capacity. In this case, the stationary probability of n customers in the system is given by

$$\pi_{n,\mu,\psi,q}^\infty = \begin{cases} \pi_{0,\mu,\psi,q}^\infty \prod_{j=0}^{n-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{1, 2, \dots, n_b\}, \\ \pi_{0,\mu,\psi,q}^\infty \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \prod_{j=0}^{n_b-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{n_b + 1, n_b + 2, \dots, \infty\}, \end{cases} \tag{12}$$

where

$$\pi_{0,\mu,\psi,q}^\infty = \left[1 + \sum_{n=1}^{n_b} \prod_{j=0}^{n-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right) + \prod_{j=0}^{n_b-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right) \sum_{n=n_b+1}^\infty \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \right]^{-1}. \tag{13}$$

With the additional stability condition $\Lambda(1 - q) < \underline{\mu}$, we can ensure that $\sum_{n=n_b+1}^\infty \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b}$ is finite. The necessary and sufficient condition for the pure threshold strategy k to be in equilibrium is $\omega^\infty(n) \leq \frac{p}{1-p} \theta_q^\infty(k)$ for all $0 \leq n \leq k - 1$ and $\omega^\infty(n) \geq \frac{p}{1-p} \theta_q^\infty(k)$ for all $k \leq n \leq n_b$, where $\omega^\infty(n) =$

$$\left(\frac{\bar{\mu}}{\mu} \right)^n \frac{\underline{\mu}}{V - \frac{n+1}{\bar{\mu}}c} - V \quad \text{and} \quad \theta_q^\infty(k) = \frac{1 + \sum_{j=1}^k \left(\frac{\Lambda}{\underline{\mu}} \right)^j + \sum_{j=k+1}^\infty (1-q)^{j-k} \left(\frac{\Lambda}{\underline{\mu}} \right)^j}{1 + \sum_{j=1}^k \left(\frac{\Lambda}{\bar{\mu}} \right)^j + \sum_{j=k+1}^\infty (1-q)^{j-k} \left(\frac{\Lambda}{\bar{\mu}} \right)^j}.$$

It follows that the existence of at least one pure threshold strategy equilibrium is guaranteed (Proposition 1). Propositions 2 and 3 also continue to hold.

8 Conclusion

In the analysis of an $M/M/1/\bar{N}$ non-customer-intensive service queue with strategic impatient and non-strategic patient customers, we find that there exists at least one pure threshold strategy equilibrium in the impatient customers’ game. With the notion of FTC behavior, we also show the possibility of multiple threshold equilibria in pure or mixed space. Moreover, we understand the effect of the proportion of impatient customers and the waiting area capacity on the likelihood ratio of a fast server to a slow server in an empty queue and the minimal pure threshold strategy equilibrium. We show that the likelihood ratio is monotonically decreasing in the proportion of impatient customers and monotonically increasing in the waiting area capacity. Further, we show that the minimal pure threshold strategy equilibrium is non-increasing in the proportion of impatient customers and non-decreasing in the waiting area capacity.

The pure threshold strategy equilibrium is also guaranteed to exist when the waiting area capacity is infinite. We then numerically analyze the effect of the proportion of impatient customers and the waiting area capacity on the expected throughput and the expected blocking rate.

Our equilibrium analysis comes with certain limitations. We did not analyze whether a non-threshold strategy equilibrium exists in this setting. Investigating the existence of non-threshold strategy equilibrium can generate further insights. Secondly, the non-strategic nature of patient customers is not exactly applicable in certain service environments. Instead, the patient customer could always join the queue and strategically decide whether to revisit the service facility or not. This is especially true for services that involve repeated visits by the customers. Studies have looked at revisit decisions in service systems, albeit on a limited scale [17,26]. This notion of patient customers making a strategic revisit decision is a promising extension to the analysis of queueing games in services.

Appendix. Proofs

Proof of Lemma 1 In the birth–death process, solving the flow balance equations yields

$$\pi_{n,\mu,\psi,q} = \begin{cases} \left([\psi_{n-1}q + (1-q)] \frac{\Lambda}{\mu} \right) \pi_{n-1,\mu,\psi,q} = \dots & n \in \{1, 2, \dots, n_b\}, \\ = \prod_{j=0}^{n-1} \left([\psi_jq + (1-q)] \frac{\Lambda}{\mu} \right) \pi_{0,\mu,\psi,q}, & \\ = \left(\frac{(1-q)\Lambda}{\mu} \right) \pi_{n-1,\mu,\psi,q} = \dots & \\ = \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \prod_{j=0}^{n_b-1} \left([\psi_jq + (1-q)] \frac{\Lambda}{\mu} \right) \pi_{0,\mu,\psi,q}, & n \in \{n_b + 1, \dots, \bar{N}\}. \end{cases}$$

Also, using the property that $\sum_{n=0}^{\bar{N}} \pi_{n,\mu,\psi,q} = 1$, we get

$$\pi_{0,\mu,\psi,q} = \left[1 + \sum_{n=1}^{n_b} \prod_{j=0}^{n-1} \left([\psi_jq + (1-q)] \frac{\Lambda}{\mu} \right) + \prod_{j=0}^{n_b-1} \left([\psi_jq + (1-q)] \frac{\Lambda}{\mu} \right) \sum_{n=n_b+1}^{\bar{N}} \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \right]^{-1}$$

and therefore,

$$\pi_{n,\mu,\psi,q} = \begin{cases} \pi_{0,\mu,\psi,q} \prod_{j=0}^{n-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{1, 2, \dots, n_b\}, \\ \pi_{0,\mu,\psi,q} \left(\frac{(1-q)\Lambda}{\mu} \right)^{n-n_b} \prod_{j=0}^{n_b-1} \left([\psi_j q + (1-q)] \frac{\Lambda}{\mu} \right), & n \in \{n_b + 1, n_b + 2, \dots, \bar{N}\}, \end{cases}$$

which completes the proof. □

Proof of Lemma 2 For the pure threshold strategy to be in equilibrium, the customer’s utility $U(n, k, q) \geq 0$ for all $0 \leq n \leq k - 1$ and $U(n, k, q) \leq 0$ for all $k \leq n \leq n_b$. Hence, it follows that, for all $0 \leq n \leq k - 1$,

$$p\pi_{n,\bar{\mu},k,q} \left(V - \frac{n+1}{\bar{\mu}}c \right) + (1-p)\pi_{n,\underline{\mu},k,q} \left(V - \frac{n+1}{\underline{\mu}}c \right) \geq 0.$$

Then,

$$\pi_{n,\underline{\mu},k,q} \left(\frac{n+1}{\underline{\mu}}c - V \right) \leq \frac{p}{1-p} \pi_{n,\bar{\mu},k,q} \left(V - \frac{n+1}{\bar{\mu}}c \right).$$

Simplifying the expression further completes the proof. □

Proof of Lemma 3 (i) Let $\underline{\rho} = \frac{\Lambda}{\underline{\mu}}$ and $\bar{\rho} = \frac{\Lambda}{\bar{\mu}}$ where $\bar{\rho} < \underline{\rho}$. We first write $\theta_q(\tilde{k}, \bar{N})$ as

$$\theta_q(\tilde{k}, \bar{N}) = \frac{\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j}{\sum_{j=0}^{\tilde{k}} \bar{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j}.$$

It is evident that $\theta_q(\tilde{k}, \bar{N})$ is independent of \tilde{k} at $q = 0$ and therefore, $\frac{d\theta_q(\tilde{k}, \bar{N})}{d\tilde{k}} = 0$.

At $q = 1$, $\theta_q(\tilde{k}, \bar{N})$ in our case is the inverse of $\Phi(\tilde{k})$ in Debo and Veeraraghavan [5]. Therefore, at $q = 1$, it is easy to see that $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \tilde{k} and $\frac{d\theta_q(\tilde{k}, \bar{N})}{d\tilde{k}} > 0$.

We are now interested in the behavior of $\theta_q(\tilde{k}, \bar{N})$ in \tilde{k} when $q \in (0, 1)$. Consider two queue lengths $z \in \{1, 2, \dots\}$ and $z - 1 \in \{0, 1, 2, \dots\}$. Our aim is to prove that $\theta_q(z, \bar{N}) - \theta_q(z - 1, \bar{N}) > 0$. In order to prove that $\theta_q(z, \bar{N}) - \theta_q(z - 1, \bar{N}) > 0$, it

suffices to show that Eq. (14) is satisfied:

$$\begin{aligned} & \left(\sum_{j=0}^z \underline{\rho}^j + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \right) \left(\sum_{j=0}^{z-1} \bar{\rho}^j + \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j \right) \\ & - \left(\sum_{j=0}^z \bar{\rho}^j + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \right) \left(\sum_{j=0}^{z-1} \underline{\rho}^j + \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j \right) > 0. \end{aligned} \tag{14}$$

When we expand the terms in the aforementioned expression, we need to obtain

$$\begin{aligned} & \left(\sum_{j=0}^z \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j \right) + \left(\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j \right) \\ & + \left(\sum_{j=0}^z \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j \right) \\ & + \left(\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j \right) > 0. \end{aligned} \tag{15}$$

Consider $\sum_{j=0}^z \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j$ from Eq. (15):

$$\begin{aligned} \sum_{j=0}^z \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j &= \left(\underline{\rho}^z + \sum_{j=0}^{z-1} \underline{\rho}^j \right) \sum_{j=0}^{z-1} \bar{\rho}^j - \left(\bar{\rho}^z + \sum_{j=0}^{z-1} \bar{\rho}^j \right) \sum_{j=0}^{z-1} \underline{\rho}^j \\ &= \underline{\rho}^z \sum_{j=0}^{z-1} \bar{\rho}^j - \bar{\rho}^z \sum_{j=0}^{z-1} \underline{\rho}^j \\ &= \underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right) \\ &= \underline{\rho}^z \bar{\rho}^z \left(\frac{1}{\bar{\rho}^z} - \frac{1}{\underline{\rho}^z} + \frac{1}{\bar{\rho}^{z-1}} - \frac{1}{\underline{\rho}^{z-1}} + \dots + \frac{1}{\bar{\rho}} - \frac{1}{\underline{\rho}} \right). \end{aligned} \tag{16}$$

We know that $\underline{\rho} > \bar{\rho} \implies \frac{1}{\underline{\rho}} < \frac{1}{\bar{\rho}} \implies \frac{1}{\underline{\rho}^z} < \frac{1}{\bar{\rho}^z}$ for all $z > 0$. It follows that

$$\sum_{j=0}^z \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j > 0. \tag{17}$$

Now consider $\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j$ from Eq. (15):

$$\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j - \sum_{j=0}^z \bar{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j = \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j$$

$$\begin{aligned}
 & - \left(\sum_{j=0}^{z-1} \underline{\rho}^j + \underline{\rho}^z \right) \left((1-q)\underline{\rho}^z + (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \right) \\
 = & \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j - (1-q)\underline{\rho}^z \sum_{j=0}^{z-1} \underline{\rho}^j - (1-q)\underline{\rho}^z \underline{\rho}^z \\
 & - (1-q) \sum_{j=0}^{z-1} \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - \underline{\rho}^z (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \\
 = & q \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j - (1-q)\underline{\rho}^z \underline{\rho}^z \left(\sum_{j=0}^{z-1} \underline{\rho}^{j-z} + 1 + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z} \right). \tag{18}
 \end{aligned}$$

Consider $\sum_{j=0}^z \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j$ from Eq. (15):

$$\begin{aligned}
 & \sum_{j=0}^z \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j \\
 = & \left(\sum_{j=0}^{z-1} \underline{\rho}^j + \underline{\rho}^z \right) \left((1-q)\underline{\rho}^z + (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \right) \\
 & - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j = (1-q)\underline{\rho}^z \sum_{j=0}^{z-1} \underline{\rho}^j + (1-q)\underline{\rho}^z \underline{\rho}^z \\
 & + (1-q) \sum_{j=0}^{z-1} \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j + (1-q)\underline{\rho}^z \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \\
 & - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j = (1-q)\underline{\rho}^z \underline{\rho}^z \left(\sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right. \\
 & \left. + 1 + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z} \right) - q \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j. \tag{19}
 \end{aligned}$$

Consider $\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j$ from Eq. (15):

$$\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j$$

$$\begin{aligned}
 &= \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \left((1-q)\bar{\rho}^z + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j \right) \\
 &\quad - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \left((1-q)\underline{\rho}^z + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j \right) \\
 &= (1-q)\bar{\rho}^z \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - (1-q)\underline{\rho}^z \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \\
 &\quad + \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j \\
 &\quad - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j \\
 &= (1-q)\bar{\rho}^z \underline{\rho}^z \left(\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z} - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^{j-z} \right) \\
 &\quad + (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \\
 &\quad - (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \\
 &= (1-q)\bar{\rho}^z \underline{\rho}^z \left(\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z} - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^{j-z} \right). \tag{20}
 \end{aligned}$$

It is apparent that the sum of $(1-q)\bar{\rho}^z \underline{\rho}^z \left(\sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z} - \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^{j-z} \right)$ from Eq. (20), $-(1-q)\underline{\rho}^z \bar{\rho}^z \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^{j-z}$ from Eq. (18) and $(1-q)\bar{\rho}^z \underline{\rho}^z \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^{j-z}$ from Eq. (19) is equal to 0.

From Eqs. (16), (18) and (19), we find the sum of $\underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right)$, $-(1-q)\underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} + 1 \right)$ and $(1-q)\bar{\rho}^z \underline{\rho}^z \left(\sum_{j=0}^{z-1} \underline{\rho}^{j-z} + 1 \right)$, which is equal to

$$\begin{aligned}
 &\underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right) + (1-q)\underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \underline{\rho}^{j-z} - \sum_{j=0}^{z-1} \bar{\rho}^{j-z} \right) \\
 &= \underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right) - (1-q)\underline{\rho}^z \bar{\rho}^z \left(-\sum_{j=0}^{z-1} \underline{\rho}^{j-z} + \sum_{j=0}^{z-1} \bar{\rho}^{j-z} \right)
 \end{aligned}$$

$$= q \underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right) > 0. \tag{21}$$

From Eq. (17), it follows that $q \underline{\rho}^z \bar{\rho}^z \left(\sum_{j=0}^{z-1} \bar{\rho}^{j-z} - \sum_{j=0}^{z-1} \underline{\rho}^{j-z} \right) > 0$ as $q \in (0, 1)$. Finally, consider the terms $q \sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \underline{\rho}^j \sum_{j=0}^{z-1} \bar{\rho}^j$ and $-q \sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \bar{\rho}^j \sum_{j=0}^{z-1} \underline{\rho}^j$ from Eqs. (18) and (19), respectively. Here, we know that $\sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \underline{\rho}^j > \sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \bar{\rho}^j$. Adding these equations to Eq. (21), we obtain

$$\left(q \underline{\rho}^z + q \sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \underline{\rho}^j \right) \sum_{j=0}^{z-1} \bar{\rho}^j - \left(q \bar{\rho}^z + q \sum_{j=z+1}^{\bar{N}} (1 - q)^{j-z} \bar{\rho}^j \right) \sum_{j=0}^{z-1} \underline{\rho}^j > 0. \tag{22}$$

From Eq. (22), we conclude that $\theta_q(\tilde{k}, \bar{N})$ is always monotonically increasing in \tilde{k} if $q > 0$.

(ii) We consider $\omega(\tilde{n})$, which is written as

$$\omega(\tilde{n}) = \left(\frac{\bar{\mu}}{\underline{\mu}} \right)^{\tilde{n}} \frac{\tilde{n} + 1}{\underline{\mu}} c - V = \left(\frac{\bar{\mu}}{\underline{\mu}} \right)^{\tilde{n}+1} \frac{\tilde{n} - \underline{n}}{\bar{n} - \tilde{n}}.$$

As $\tilde{n} \in [\underline{n}, \bar{n}]$ and $\bar{\mu} > \underline{\mu}$, it is straightforward to observe that $\left(\frac{\bar{\mu}}{\underline{\mu}} \right)^{\tilde{n}+1}$ is increasing in \tilde{n} . Also, $\tilde{n} - \underline{n}$ is increasing in \tilde{n} and $\bar{n} - \tilde{n}$ is decreasing in \tilde{n} . We know that $\omega(\tilde{n}) \geq 0$ for $\tilde{n} \in [\underline{n}, \bar{n}]$. As the function $\omega(\tilde{n})$ is comprised of a product of two positive and increasing functions of \tilde{n} divided by a decreasing function of \tilde{n} , we infer that $\omega(\tilde{n})$ is monotonically increasing in \tilde{n} . \square

Proof of Lemma 4 Let $\theta_q(\tilde{k}, \bar{N}) = \frac{\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1 - q)^{j-\tilde{k}} \underline{\rho}^j}{\sum_{j=0}^{\tilde{k}} \bar{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1 - q)^{j-\tilde{k}} \bar{\rho}^j}$. We consider

q and $q - \epsilon$ and then aim to show that $\theta_{q-\epsilon}(\tilde{k}, \bar{N}) - \theta_q(\tilde{k}, \bar{N}) > 0$.

We can denote $\sum_{j=\tilde{k}+1}^{\bar{N}} (1 - q + \epsilon)^{j-\tilde{k}} \underline{\rho}^j = \alpha \sum_{j=\tilde{k}+1}^{\bar{N}} (1 - q)^{j-\tilde{k}} \underline{\rho}^j$, where $\alpha > 1$. Then, in order to prove that $\theta_{q-\epsilon}(\tilde{k}, \bar{N}) - \theta_q(\tilde{k}, \bar{N}) > 0$, it is enough to consider and prove Eq. (23):

$$\begin{aligned} & \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \alpha \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \left(\sum_{j=0}^{\tilde{k}} \bar{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j \right) \\ & - \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \left(\sum_{j=0}^{\tilde{k}} \bar{\rho}^j + \alpha \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j \right) > 0. \end{aligned} \tag{23}$$

From Eq. (23), we can easily show that $\sum_{j=0}^{\tilde{k}} \underline{\rho}^j \sum_{j=0}^{\tilde{k}} \bar{\rho}^j - \sum_{j=0}^{\tilde{k}} \underline{\rho}^j \sum_{j=0}^{\tilde{k}} \bar{\rho}^j = 0$ and $\alpha \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j - \alpha \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j = 0$. The remaining terms satisfy the following conditions: $\sum_{j=0}^{\tilde{k}} \bar{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j (\alpha - 1) > 0$ and $\sum_{j=0}^{\tilde{k}} \underline{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j (1 - \alpha) < 0$.

To show that $\theta_{q-\epsilon}(\tilde{k}, \bar{N}) - \theta_q(\tilde{k}, \bar{N}) > 0$, it is sufficient to show for all $\tilde{k} \in \{0, 1, 2, \dots\}$ that

$$\sum_{j=0}^{\tilde{k}} \bar{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j - \sum_{j=0}^{\tilde{k}} \underline{\rho}^j \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \bar{\rho}^j > 0. \tag{24}$$

We will use the principle of mathematical induction to prove this. At $\tilde{k} = 0$, Eq. (24) is simplified to $\sum_{j=1}^{\bar{N}} (1-q)^{j-0} \underline{\rho}^j - \sum_{j=1}^{\bar{N}} (1-q)^{j-0} \bar{\rho}^j > 0$ ($\because \underline{\rho} > \bar{\rho}$). We now assume that Eq. (24) is satisfied for $\tilde{k} = z$ and show that it is true for $\tilde{k} = z + 1$. At $\tilde{k} = z + 1$, Eq. (24) becomes

$$\begin{aligned} & \sum_{j=0}^{z+1} \bar{\rho}^j \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z+1} \underline{\rho}^j - \sum_{j=0}^{z+1} \underline{\rho}^j \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z+1} \bar{\rho}^j \\ & = \bar{\rho}^z (1-q) \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j + (1-q) \sum_{j=0}^z \bar{\rho}^j \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j \\ & \quad - \underline{\rho}^z (1-q) \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j - (1-q) \sum_{j=0}^z \underline{\rho}^j \sum_{j=z+2}^{\bar{N}} (1-q)^{j-z} \bar{\rho}^j \\ & = \bar{\rho}^z (1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - (1-q)^2 \bar{\rho}^z \underline{\rho}^{z+1} \\ & \quad + (1-q) \sum_{j=0}^z \bar{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - (1-q)^2 \sum_{j=0}^z \bar{\rho}^j \underline{\rho}^{z+1} \end{aligned}$$

$$\begin{aligned}
 & - \underline{\rho}^z(1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j + \underline{\rho}^z(1-q)^2 \underline{\rho}^{z+1} \\
 & - (1-q) \sum_{j=0}^z \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j + (1-q)^2 \sum_{j=0}^z \underline{\rho}^j \underline{\rho}^{z+1}. \tag{25}
 \end{aligned}$$

From Eq. (25), we infer that $-(1-q)^2 \underline{\rho}^z \underline{\rho}^{z+1} + \underline{\rho}^z(1-q)^2 \underline{\rho}^{z+1} < 0$. However, $\underline{\rho}^z(1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - \underline{\rho}^z(1-q) \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j > |-(1-q)^2 \underline{\rho}^z \underline{\rho}^{z+1} + \underline{\rho}^z(1-q)^2 \underline{\rho}^{z+1}| > 0$. Similarly, $-(1-q)^2 \sum_{j=0}^z \underline{\rho}^j \underline{\rho}^{z+1} + (1-q)^2 \sum_{j=0}^z \underline{\rho}^j \underline{\rho}^{z+1} < 0$. However, $(1-q) \sum_{j=0}^z \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j - (1-q) \sum_{j=0}^z \underline{\rho}^j \sum_{j=z+1}^{\bar{N}} (1-q)^{j-z} \underline{\rho}^j > |-(1-q)^2 \sum_{j=0}^z \underline{\rho}^j \underline{\rho}^{z+1} + (1-q)^2 \sum_{j=0}^z \underline{\rho}^j \underline{\rho}^{z+1}| > 0$. We now observe that Eq. (24) is satisfied for $\tilde{k} = z + 1$. Hence, $\theta_q(\tilde{k}, \bar{N})$ is monotonically decreasing in q . \square

Proof of Lemma 5 Let $\theta_q(\tilde{k}, \bar{N}) = \frac{\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j}{\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j}$. We aim to show that $\theta_q(\tilde{k}, \bar{N} + 1) - \theta_q(\tilde{k}, \bar{N}) > 0$ as presented in Equation (26):

$$\begin{aligned}
 & \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}+1} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \\
 & - \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}+1} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) > 0. \tag{26}
 \end{aligned}$$

Simplifying and rearranging the terms,

$$\begin{aligned}
 & \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j + (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \right) \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \\
 & - \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \right) \left(\sum_{j=0}^{\tilde{k}} \underline{\rho}^j + \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j + (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \right) \\
 & = (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \sum_{j=0}^{\tilde{k}} \underline{\rho}^j + (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j \\
 & - (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \sum_{j=0}^{\tilde{k}} \underline{\rho}^j - (1-q)^{\bar{N}+1-\tilde{k}} \underline{\rho}^{\bar{N}+1} \sum_{j=\tilde{k}+1}^{\bar{N}} (1-q)^{j-\tilde{k}} \underline{\rho}^j
 \end{aligned}$$

$$\begin{aligned}
 &= (1 - q)^{\bar{N}+1-\tilde{k}} \bar{\rho}^{\bar{N}+1} \underline{\rho}^{\bar{N}+1} \left(\sum_{j=0}^{\tilde{k}} \bar{\rho}^{j-\bar{N}-1} - \sum_{j=0}^{\tilde{k}} \underline{\rho}^{j-\bar{N}-1} \right) \\
 &\quad + (1 - q)^{\bar{N}+1-2\tilde{k}} \bar{\rho}^{\bar{N}+1} \underline{\rho}^{\bar{N}+1} \left(\sum_{j=0}^{\tilde{k}} ((1 - q)\bar{\rho})^{j-\bar{N}-1} - \sum_{j=0}^{\tilde{k}} ((1 - q)\underline{\rho})^{j-\bar{N}-1} \right).
 \end{aligned}
 \tag{27}$$

We know that $\underline{\rho} > \bar{\rho} \implies \frac{1}{\underline{\rho}} < \frac{1}{\bar{\rho}} \implies \frac{1}{\underline{\rho}^z} < \frac{1}{\bar{\rho}^z}$ for all $z > 0$ and $\tilde{k} < \bar{N} + 1$. It follows that $(1 - q)^{\bar{N}+1-\tilde{k}} \bar{\rho}^{\bar{N}+1} \underline{\rho}^{\bar{N}+1} \left(\sum_{j=0}^{\tilde{k}} \bar{\rho}^{j-\bar{N}-1} - \sum_{j=0}^{\tilde{k}} \underline{\rho}^{j-\bar{N}-1} \right) + (1 - q)^{\bar{N}+1-2\tilde{k}} \bar{\rho}^{\bar{N}+1} \underline{\rho}^{\bar{N}+1} \left(\sum_{j=0}^{\tilde{k}} ((1 - q)\bar{\rho})^{j-\bar{N}-1} - \sum_{j=0}^{\tilde{k}} ((1 - q)\underline{\rho})^{j-\bar{N}-1} \right) > 0$. Hence, $\theta_q(\tilde{k}, \bar{N})$ is monotonically increasing in \bar{N} . □

Proof of Proposition 4 From Lemma 5, we know that $\theta_q(k, \bar{N}_1) > \theta_q(k, \bar{N}_2)$ when $\bar{N}_1 > \bar{N}_2$. It follows that $b_q(k, \bar{N}_1) \geq b_q(k, \bar{N}_2)$ for a given k . As a pure threshold strategy equilibrium is guaranteed to exist (Proposition 1), let the equilibrium for \bar{N}_2 occur at k , i.e., $b_q(k, \bar{N}_2) = k$. If multiple equilibria exist for \bar{N}_2 , consider the highest pure threshold strategy equilibrium. Now, as $b_q(k, \bar{N}_1) \geq b_q(k, \bar{N}_2)$, we can look at two cases:

Case 1. If $b_q(k, \bar{N}_1) = b_q(k, \bar{N}_2)$, then the equilibrium for \bar{N}_1 also occurs at k .

Case 2. If $b_q(k, \bar{N}_1) > b_q(k, \bar{N}_2)$, the equilibrium for \bar{N}_1 does not occur at k . From FTC behavior, we know that $b_q(k + 1, \bar{N}_1) \geq b_q(k, \bar{N}_1)$. Now, if $b_q(k + 1, \bar{N}_1) = k + 1$, the pure threshold strategy equilibrium for \bar{N}_1 occurs at $k + 1$. Otherwise, there is no pure threshold strategy equilibrium at $k + 1$ for \bar{N}_1 . Then, we can consider $k + 2, k + 3, \dots, \lceil \bar{n} \rceil$. In general, $b_q(k + \varphi, \bar{N}_1)$ is bounded above by $\lceil \bar{n} \rceil$ as $\lim_{n \rightarrow \bar{n}} \omega(n) = +\infty$. As $b_q(k + \varphi + 1, \bar{N}_1) \geq b_q(k + \varphi, \bar{N}_1)$, we find that $b_q(k + \varphi + 1, \bar{N}_1) = \lceil \bar{n} \rceil$ if $b_q(k + \varphi, \bar{N}_1) = \lceil \bar{n} \rceil$. Therefore, if there exists no pure threshold strategy equilibrium from $k + 1$ (i.e., $b_q(k + 1, \bar{N}_1) \neq k + 1$) to $\lceil \bar{n} \rceil - 1$ (i.e., $b_q(\lceil \bar{n} \rceil - 1, \bar{N}_1) \neq \lceil \bar{n} \rceil - 1$), there will always exist a pure threshold strategy equilibrium for \bar{N}_1 at $\lceil \bar{n} \rceil$ (i.e., $b_q(\lceil \bar{n} \rceil, \bar{N}_1) = \lceil \bar{n} \rceil$). □

References

1. Anand, K.S., Paç, M.F.: Quality–speed conundrum: trade-offs in customer-intensive services. *Manag. Sci.* **57**(1), 40–56 (2011)
2. Boudali, O., Economou, A.: The effect of catastrophes on the strategic customer behavior in queueing systems. *Naval Res. Logist.* **60**(7), 571–587 (2013)
3. Debo, L.G., Toktay, L.B., Van Wassenhove, L.N.: Queuing for expert services. *Manag. Sci.* **54**(8), 1497–1512 (2008)
4. Debo, L.G., Parlour, C., Rajan, U.: Signaling quality via queues. *Manag. Sci.* **58**(5), 876–891 (2012)
5. Debo, L., Veeraraghavan, S.: Equilibrium in queues under unknown service times and service value. *Oper. Res.* **62**(1), 38–57 (2014)
6. Economou, A., Kanta, S.: Optimal balking strategies and pricing for the single server Markovian queue with compartmented waiting space. *Queueing Syst.* **59**(3–4), 237 (2008)

7. Economou, A., Kanta, S.: Equilibrium customer strategies and social-profit maximization in the single-server constant retrial queue. *Naval Res. Logist.* **58**(2), 107–122 (2011)
8. Guo, P., Zipkin, P.: The effects of the availability of waiting-time information on a balking queue. *Eur. J. Oper. Res.* **198**(1), 199–209 (2009)
9. Hassin, R., Haviv, M.: Equilibrium threshold strategies: the case of queues with priorities. *Oper. Res.* **45**(6), 966–973 (1997)
10. Hassin, R., Haviv, M.: To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems, vol. 59. Springer, Berlin (2003)
11. Hassin, R.: *Rational Queueing*. CRC Press, Boca Raton (2016)
12. Hassin, R., Roet-Green, R.: The impact of inspection cost on equilibrium, revenue, and social welfare in a single-server queue. *Oper. Res.* **65**(3), 804–820 (2017)
13. Haviv, M.: When to arrive at a queue with tardiness costs? *Perform. Eval.* **70**(6), 387–399 (2013)
14. Iravani, F., Balcioglu, B.: On priority queues with impatient customers. *Queueing Syst.* **58**(4), 239 (2008)
15. Juneja, S., Shimkin, N.: The concert queueing game: strategic arrivals with waiting and tardiness costs. *Queueing Syst.* **74**(4), 369–402 (2013)
16. Kakutani, S.: A generalization of Brouwer's fixed point theorem. *Duke Math. J.* **8**(3), 457–459 (1941)
17. Kostami, V., Rajagopalan, S.: Speed-quality trade-offs in a dynamic model. *Manuf. Serv. Oper. Manag.* **16**(1), 104–118 (2013)
18. Kulkarni, V.G.: On queueing systems by retrials. *J. Appl. Probab.* **20**(2), 380–389 (1983)
19. Li, L., Wang, J., Zhang, F.: Equilibrium customer strategies in markovian queues with partial breakdowns. *Comput. Ind. Eng.* **66**(4), 751–757 (2013)
20. Liu, Y., Cooper, W.L.: Optimal dynamic pricing with patient customers. *Oper. Res.* **63**(6), 1307–1319 (2015)
21. Liu, Y., Shang, W.: Follow the crowd to avoid congestion with unknown service rate. In: The 10th POMS-HK International Conference 2019 : Operations Excellence for a Better World ; Conference date: 05-01-2019 Through 06-01-2019. <http://www.cb.cityu.edu.hk/ms/pomshk2019/conferenceprogram.htm>
22. Nair, S.: 'Virar crowd' loses unity on packed local trains. *The Times of India*. <https://timesofindia.indiatimes.com/city/mumbai/Virar-crowd-loses-unity-on-packed-local-trains/articleshow/53440493.cms> (2016). Accessed 26 September 2019
23. Naor, P.: The regulation of queue size by levying tolls. *Econom. J. Econom. Soc.* **37**, 15–24 (1969)
24. Sumita, U., Masuda, Y., Yamakawa, S.: Optimal internal pricing and capacity planning for service facility with finite buffer. *Eur. J. Oper. Res.* **128**(1), 192–205 (2001)
25. Wolff, R.W.: Poisson arrivals see time averages. *Oper. Res.* **30**(2), 223–231 (1982)
26. Xiao, G., Dong, M., Li, J., Sun, L.: Scheduling routine and call-in clinical appointments with revisits. *Int. J. Prod. Res.* **55**(6), 1767–1779 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.