Submitted to *Operations Research*
manuscript (Please, provide the mansucript number!)

Bounded Rationality in Service Systems

Tingliang Huang
Kellogg School of Management, Northwestern University, Evanston, IL, tingliang-huang@kellogg.northwestern.edu

Gad Allon
Kellogg School of Management, Northwestern University, Evanston, IL, g-allon@kellogg.northwestern.edu

Achal Bassamboo
Kellogg School of Management, Northwestern University, Evanston, IL, a-bassamboo@kellogg.northwestern.edu

The traditional economics and queueing literature typically assume that customers are fully rational. In contrast, in this paper, we study canonical service models with boundedly rational customers. We capture bounded rationality using a framework in which better decisions are made more often, while the best decision needs not always be made. We investigate the impact of bounded rationality on social welfare and revenue of a profit maximizing firm when the queue is visible and not visible to the customers. For invisible queues, from the firm's perspective, higher irrationality always leads to higher optimal prices and higher revenue when customers are sufficiently irrational. From the social planner's perspective, there may be strictly positive social welfare losses when customers are sufficiently irrational. For visible queues with a fixed price, we prove that a little bit of irrationality can lead to strict social welfare improvement, and we provide a simple inequality under which this improvement happens. With the optimal prices, however, irrationality always decreases social welfare, and a little bit of irrationality always results in revenue losses.

Key words: Behavioral operations, bounded rationality, consumer information, customer behavior, economics, product quality, queueing systems, service operations, social welfare

1. Introduction and Summary

When a customer calls a call center or goes to a fast food restaurant, a café or an ATM, and has to queue for service, does he always *accurately* and *perfectly* calculate the benefits and costs of joining before making his decisions? The traditional economics and queueing literature have assumed that he does, while anecdotal evidence and experimental studies point to the contrary. In this paper, we study queueing or service systems without making this "perfect rationality" assumption on the part of customers.

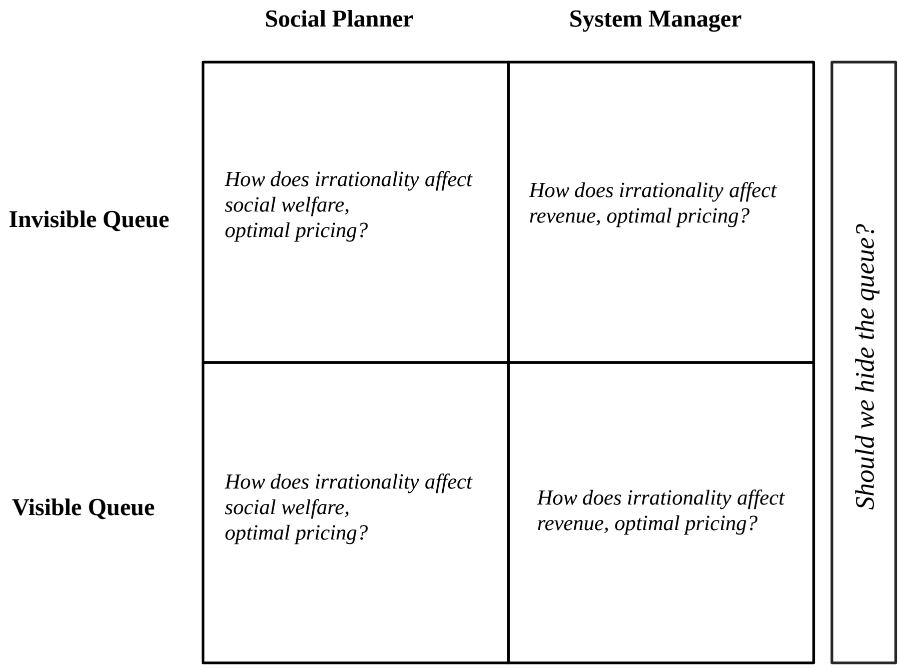
Naor (1969) seems to be the first to realize the fact that customers are decision makers. Naor

(1969) and subsequent researchers following his work assume customers are self-interested, can *perfectly* optimize their own utilities, and make decisions *without* any mistakes. Assuming all customers are fully rational, Naor (1969) shows that self-interested customers would join a more congested system than what the the social planner prescribes, and proposes "levying tolls" (i.e., pricing) as a way to maximize social welfare. In Naor's model, customers are assumed to be able to compute with great precision the expected utility they are about to obtain from making a decision about whether to join or renege. Furthermore, customers are assumed to be perfectly rational and prefer joining even if the positive expected utility is negligible. One may ask, Are customers fully rational? Specifically, does a customer join a queueing system even if his positive expected utility is negligibly small? Ariely (2009) claims that irrationality is the real invisible hand that drives human decision making. Indeed, while we are not aware of any empirical evidence that people are fully rational, there is abundant empirical evidence that people are boundedly rational (the related literature review follows in the next section). In this work, we study the effects and implications of the bounded rationality in canonical queueing or service systems.

While customers have traditionally been assumed to perfectly maximize their expected utility, we choose to use the logit choice model, which is derived from the classical quantal choice theory (Luce 1959). This model captures the consistency property: While the best option is not always chosen, better options are chosen more often (Thurstone 1927, Luce 1959, Blume 1993, McKelvey and Palfrey 1995, and Su 2008). Further, this model parameterizes the customers' *irrationality level* by β, which allows us to have a continuum of levels of bounded rationality including the two extremes: a) *full rationality* where customers are perfect utility-maximizers; and b) *full irrationality* where each customer randomizes with equal probabilities among all choices available.

The research setting of this paper, along with our research questions, is depicted in Figure 1. We study invisible queues (such as a call center) and visible queues (such as a fast food restaurant, a café or an ATM) separately, from both a social planner's perspective and a revenue-maximizing firm's perspective. We investigate the impacts of the bounded rationality on the part of customers.

Figure 1 Research Questions



Finally, we are interested in whether the queue should be hidden when price regulation is impossible (as argued in Hassin 1986).

Next, we summarize our main results:

Invisible Queue. In the setting in which the queue is invisible, we prove that there always exists a unique equilibrium for any price and finite irrationality level. Furthermore, in equilibrium, each customer uses a non-degenerate probability to join the queue. We show that, for any fixed price, the revenue function is either decreasing or increasing in the irrationality level depending on whether the equilibrium joining probability under full rationality is above one half or not. At the same time, for any fixed irrationality level, the revenue function is *unimodal* in price, and thus there exists a unique price to maximize the revenue. We also show that this revenue-maximizing price and the optimal revenue are both strictly increasing in the irrationality level when customers are sufficiently irrational. Therefore, the revenue-maximizing firm can exploit the bounded rationality when customers are sufficiently irrational.

In the same system, we prove that the social welfare function is *unimodal* in the irrationality level for a fixed price. Moreover, surprisingly, the social welfare is strictly increasing in the irrationality level under certain conditions. We also show that the social welfare function is *unimodal* in price for a fixed irrationality level, and we derive the unique socially optimal price in closed form. The optimal social welfare (i.e., the social welfare under optimal pricing) is independent of both irrationality level and arrival rate if customers are not too irrational. Therefore, the social planner can always correct for the irrationality by charging an appropriate price as long as customers are not too irrational. In contrast, when customers are too irrational, there exist social welfare losses.

Visible Queue. In the setting in which the queue is visible to the customers and the price is fixed, we show that this system is always stable as long as customers are not fully irrational. We provide a rigorous characterization showing whether and when the social welfare increases when customers are slightly irrational. To better understand the result, recall that Naor (1969) shows that when customers are fully rational, self-interested customers join a more congested system than the socially optimal one. Boundedly rational customers err on both sides, joining a more congested system and balking when congestion is low. While the former is detrimental to the social welfare, the latter can be beneficial. The social welfare can thus improve depending on which one of these effects dominates. We provide a simple characterization of this dichotomy. While one can characterize how the social welfare behaves when customers are slightly irrational, we find that it is not necessarily monotone or unimodal with respect to the irrationality level, which stands in contrast to the case of invisible queues.

We also study the settings when prices are charged optimally (either by a social planner or a revenue-maximizing firm). We prove that bounded rationality always results in social welfare losses. In other words, when an arbitrary price is charged by the social planner, the social welfare can increase when customers are slightly irrational, yet it can only decrease when the optimal prices are charged. Thus, the social planer cannot correct for the irrationality when the queue is visible to customers, which is in contrast to the case of invisible queues. We also show that a little bit of irrationality always makes revenue-collecting less profitable, which again stands in contrast to the case of invisible queues.

Finally, we numerically demonstrate that bounded rationality has non-trivial impacts on the policy recommendations such as whether or not to hide the queue.

There are two main contributions of the paper. First, we seem to be among the first to apply the quantal choice framework to canonical queueing settings. We take a descriptive approach and consider bounded rationality, in the sense that decision makers are prone to errors and biases, instead of taking a normative approach. Second, within this framework, we show the impact of not accounting for bounded rationality of customers on the ability to manage queues.

The remainder of this paper is organized as follows. The literature is reviewed in §2. §3 presents the model with the invisible queue. We study the model with the visible queue in §4. In §5, we study whether the social planner should reveal (or enforce the revenue-maximizing firm to reveal) the real-time congestion information of the queue. We provide conclusions in the last section. Proofs of all the results are relegated to the Appendix.

2. Literature Review

Our study is related to several branches of the literature: bounded rationality in economics, economics of queues and behavioral operations.

Bounded rationality in economics: Traditional economic theory postulates that decision makers are "rational," i.e., they have sufficient abilities to do perfect optimization in their choices. Simon (1955) seems to be the first to propose an alternative way to model decision-making behavior: rather than optimizing perfectly, agents search over the alternatives until they find "satisfactory" solutions. Simon (1957) coins the term "bounded rationality" to describe such human behavior. A growing empirical evidence calls into question the full rationality (the utility maximization paradigm, for instance) on the part of the decision makers. For a description of systematic errors made by experimental subjects, see Arkes and Hammond (1985), Hogarth (1980), Kahneman, Slovic, and Tversky (1981), Nisbett and Ross (1980), and the survey papers by Payne, Bettman, and Johnson (1992) and by Pitz and Sachs (1984). Tversky and Kahnemann (1974) show that people rely on a limited number of heuristic principles which in general are useful but sometimes lead to severe and systematic errors. On the basis of the evidence, Conlisk (1996) offers four convincing reasons for incorporating bounded rationality in economic models. Geigerenzer and Selten (2001) adopt heuristics or rules of thumb to model bounded rationality. McKelvey and Palfrey (1995) develop a new equilibrium concept quantal response equilibrium (QRE), which incorporates the idea of bounded rationality, modeled as probabilistic choice, into game theory. This concept is adopted in this paper. Fehr and Schmidt (2006) provide a chapter on how the self-interest hypothesis that assumes that material self-interest exclusively motivates all people is unrealistic. Rather, concerns for altruism, fairness, and reciprocity strongly motivate many people. We refer readers to Rubinstein (1998) for a comprehensive treatment of bounded rationality.

There has recently been a growing literature in different settings explicitly incorporating bounded rationality. Tirole (2007) develops a model of bounded rationality and examines its consequences for contractual design. Bajari and Hortacsu (2001, 2005) study bounded rationality in auction models. Basov (2009) studies a monopolistic screening problem with boundedly rational consumers. Tharakunnel and Bhattacharyya (2009) consider single-leader-multiple-follower games with boundedly rational agents.

Economics of queues: Naor (1969) studies the economics of queueing systems when customers are fully rational. Thereafter, extensive research has been conducted in this direction. Yechiali (1971, 1972) extends Naor's model to allow for $GI/M/1$ queues. Knudsen (1972) extends Naor's model to allow for a multi-server queueing system in which arriving customers' net benefits are heterogeneous. Lippman and Stidham (1977) extend the Naor model to the finite-horizon and discounted cases showing that, in these settings, the economic notion of an external effect has a precise quantitative interpretation. Hassin (1986) considers a revenue maximizing server who has the opportunity to suppress information on actual queue length, leaving customers to decide on joining the queue on the basis of the known distribution of waiting times. He shows that it may be, but is not always, socially optimal to prevent suppression and that it is never optimal to encourage suppression when the revenue maximizer prefers to reveal the queue length. For other extensions, see Van Mieghem (2000), Hassin and Haviv (2003), Afèche (2004), Armony and Maglaras (2004a, b), Lariviere and Van Mieghem (2004), Allon and Bassamboo (2008), Hsu et al. (2009), Haviv (2009), Çil et al. (2009), and Hassin (2009) for a comprehensive literature review. Although various models along this line are studied, one common theme in this literature is that full rationality is always assumed.

Behavioral operations: Gino and Pisano (2008) survey the literature on modeling bounded rationality in economics, finance, and marketing and argue that operations management scholars should incorporate departures from the rationality assumption into their models and theories. There is an emerging literature on behavioral operations management, and we refer the readers to Bendoly et al. (2006) and Bendoly et al. (2008) for this stream of research. We point out two papers that are closely related to our work: Su (2008) is among the first papers to study bounded rationality in operational settings. He applies the logit choice framework to the classic newsvendor model and characterizes the ordering decisions made by a boundedly rational decision maker. He identifies systematic biases and investigates the impact of these biases on several operational settings. We apply a similar framework in this paper, but rather to a queueing setting. Recently, Plambeck and Wang (2010) study implications of hyperbolic discounting in service systems. Although the research setting (i.e., service systems) is similar, the research focus and approach are quite different. In their paper, customers lack the self control to undergo an unpleasant experience that would be in their long-run self interest, which is modeled by psychologists in terms of a hyperbolic discount rate for utility. In our paper, we model bounded rationality by incorporating *stochastic elements* into customers' decisions using the logit choice framework.

3. The Invisible Queue

3.1. The Model

Consider a single-server queueing system. Potential customers (who are boundedly rational) arrive to this system according to a Poisson process with rate λ. Since customers cannot observe the state of the system, they have to make a decision a priori whether to arrive to the queue or not. A customer that does not arrive to the queue receives zero utility. If a customer decides to arrive, he pays a price p and receives a reward of R on completion of service, $R > 0$. He also incurs a cost of C per unit of time while staying in the system (either waiting or being served), $C > 0$. Service times are assumed to be independently, identically, and exponentially distributed with mean $\frac{1}{\mu}$. Customers are served on a first-come first-served basis.

In investigating the system, we are first interested in each customer's strategy. Each customer decides whether or not to join the queue with probability $\varphi(p,\beta)$, $\varphi(p,\beta) \in [0,1]$, where β measures the *irrationality level*, i.e., the extent of bounded rationality of the customer. A customer's net benefit or utility of joining is $R - p - \frac{C}{(\mu - \varphi(p,\beta)\lambda)^+}$, where we used the fact that the *thinning* of a Poisson process with arrival rate λ is still a Poisson process with rate $\varphi(p,\beta)\lambda$ and $\frac{C}{(\mu - \varphi(p,\beta)\lambda)^+}$ is the customer's expected waiting cost. The arrival rate $\varphi(p,\beta)\lambda$ of the queue will be referred to as effective demand. As discussed in the literature review, each customer's joining probability is defined as a logit probability in terms of their expected utility of joining and irrationality level. Specifically, we define the equilibrium of this queueing system as follows.

Definition 1. **(Equilibrium Joining Probability)**. We say that $\varphi(p,\beta)$ is an equilibrium joining probability if it satisfies the following

$$\varphi(p,\beta) = \frac{e^{\frac{R-p-\frac{C}{(\mu-\varphi(p,\beta)\lambda)^+}}{\beta}}}{1 + e^{\frac{R-p-\frac{C}{(\mu-\varphi(p,\beta)\lambda)^+}}{\beta}}}, \tag{1}$$

for $\beta > 0$, and

$$\varphi(p,0) = \min\{\varphi_0, 1\},$$

where φ_0 satisfies

$$R - p - \frac{C}{\mu - \varphi_0\lambda} = 0, \tag{2}$$

for $\beta = 0$.

When $\beta > 0$, equation (1) yields a fixed-point problem given that the logit expression in the RHS includes the equilibrium joining probability (i.e., the LHS).

When $\beta = 0$, i.e., customers are fully rational, then the definition is precisely Hassin (1986)'s equilibrium condition (equation (4.1) on Page 1189). It is possible that there is no $\varphi_0 \in [0,1]$ satisfying equation (2) and the actual arrival rate then is λ since, even if all customers decide to join, each customer's expected utility is still strictly positive. According to this definition, we have $\varphi(p,0) = \min\{\frac{\mu}{\lambda} - \frac{C}{\lambda(R-p)}, 1\}$.

Next, we investigate whether an equilibrium always exists. Proposition 1 shows that there always exists a unique equilibrium.

PROPOSITION 1. *There always exists a unique equilibrium for the invisible queue, for any finite price p and irrationality level $\beta > 0$.*

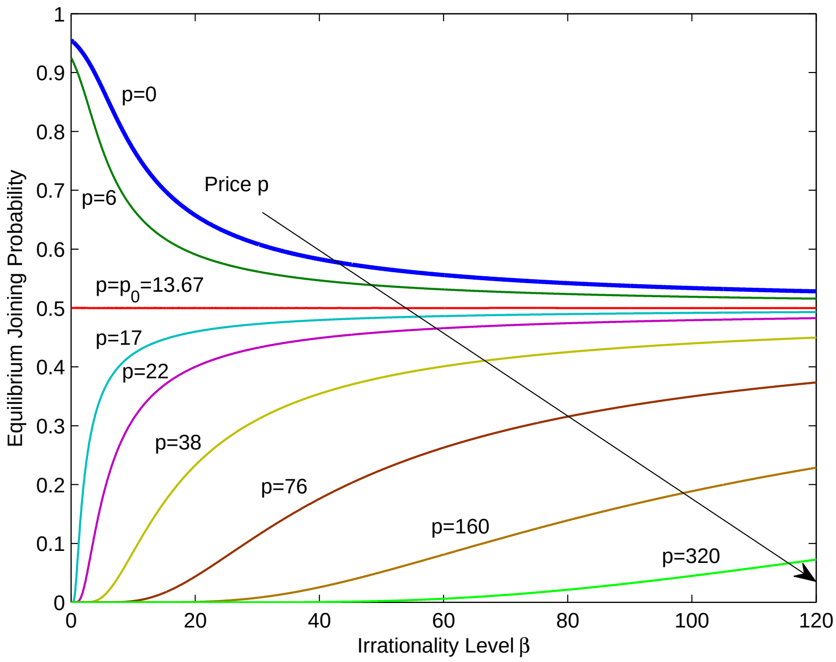
We are now interested in how the (unique) equilibrium $\varphi(p,\beta)$ behaves as a function of the price p and the irrationality level β. For convenience, we let $\overline{p} := R - \frac{2C}{2\mu - \lambda}$ denote the price under which each customer receives exactly zero utility so that the equilibrium joining probability is half *regardless of* the irrationality level. The following proposition characterizes the equilibrium joining probability.

PROPOSITION 2. *(i) If $p < \overline{p}$, equilibrium joining probability $\varphi(p,\beta)$ is strictly decreasing in β.*
(ii) If $p > \overline{p}$, equilibrium joining probability $\varphi(p,\beta)$ is strictly increasing in β.
(iii) If $p = \overline{p}$, equilibrium joining probability $\varphi(p,\beta) = \frac{1}{2}$ for any β.
(iv) For any fixed β, equilibrium joining probability $\varphi(p,\beta)$ is strictly decreasing in p.

We offer the following intuition: when the price is so low that each customer receives strictly positive utility, the initial joining probability is above half. As the irrationality level increases, better decisions are made less often, and thus the joining probability decreases as customers are more irrational. Interestingly, if the price is set so that each customer receives exactly zero utility in equilibrium, then increasing the irrationality level has no effect on the joining probability since each customer randomizes with equal probabilities regardless of the irrationality level.

It is intuitively clear that $\varphi(p,\beta)$ is strictly decreasing in price p by equality (1), i.e., a larger price always results in a lower joining probability, regardless of the irrationality level, which is

Figure 2 Equilibrium Joining Probability versus Irrationality Level for Different Prices ($R = 15, C = 2, \mu = 3, \lambda = 3$.)



basically the "law of demand" in this service setting.

We illustrate this proposition by a numerical study, where we set the parameters $R = 15$, $C = 2$, $\mu = 3$, and $\lambda = 3$. For any given price p, we numerically solve the equilibrium joining probability $\varphi(p, \beta)$ for different irrationality levels β. We depict our findings in Figure 2. Figure 2 also confirms the behavior with respect to price changes.

3.2. Revenue Maximization

In the previous section, our attention has been focused on the system equilibrium. In this section, we focus our attention on the revenue generated from such a system. In this sense, we are looking from a revenue-maximizing firm's perspective. The firm's objective is to choose a price p to maximize the expected revenue $\Pi^I(p, \beta) := p\varphi(p, \beta)\lambda$, where $\varphi(p, \beta)\lambda$ is the effective demand rate to the system. Note that we normalize the cost of serving customers to zero without loss of generality.

To investigate the firm's revenue maximization problem, we first study the behavior of the revenue $\Pi^I(p,\beta)$ as a function of the price p and the irrational level β. It is easy to see from Proposition 2 how the revenue function behaves as a function of the irrationality level β for any fixed price p, since $\Pi^I(p,\beta)$ is simply a linear transformation of $\varphi(p,\beta)$.

We next investigate how the revenue $\Pi^I(p,\beta)$ behaves as a function of the price p, for any fixed irrationality level β. We prove that $\Pi^I(p,\beta)$ is *unimodal* in p. Hence, there exists a unique revenue-maximizing price $p^*(\beta)$. The firm faces the following tradeoff when determining the optimal price $p^*(\beta)$: Higher prices bring more revenue if the equilibrium joining probability were unchanged, however, higher prices actually induce strictly lower joining probabilities, as already shown in Proposition 2. Hence, $p^*(\beta)$ is the price that balances the tradeoff.

Further, we show that the revenue-maximizing price $p^*(\beta)$ strictly increases in the irrationality level β under certain conditions.

To state Proposition 3 that makes the claim rigorous, we denote $\beta_0 := \frac{1}{2}R - \frac{2\mu C}{(2\mu-\lambda)^2}$, which is the irrationality level at which the optimal price $p^*(\beta_0) = \overline{p}$.

PROPOSITION 3. *(i) For any fixed irrationality level β, $\Pi^I(p,\beta)$ is unimodal in p, and thus there exists a unique price $p^*(\beta)$ that maximizes $\Pi^I(p,\beta)$.*
(ii) The optimal price $p^*(\beta)$ is strictly increasing in the irrationality level β for $\beta \in [\max\{\beta_0, 0\}, \infty)$.

From this proposition, we obtain that the revenue-maximizing price $p^*(\beta)$ is monotonically increasing in β if $R \le \frac{4C\mu}{(2\mu-\lambda)^2}$.

When the optimal price induces each customer to receive strictly negative utility in equilibrium, a higher irrationality level would induce the firm to increase its price. The reason is that higher irrationality leads to higher joining probabilities for a fixed price in this case. Hence, when the optimal price is "already high" (so that each customer receives strictly negative utility in equilibrium), then increasing the irrationality level leads to "even higher" optimal prices. However, when the optimal price is low so that each customer receives strictly positive utility, then increasing the irrationality level *can* lead to lower optimal prices, where the firm's tradeoff is about *the benefit of higher prices* versus the *loss of lower joining probabilities.* Proposition 3 above partially characterizes which one of these effects dominates.

We are now ready to state the result on the effect of the irrationality level on the *optimal* revenue $\Pi^I(p^*(\beta),\beta)$. Using the envelope theorem, we obtain the following immediate corollary to Proposition 2.

COROLLARY 1. *(i) If $p^*(\beta) > \overline{p}$, then optimal revenue $\Pi^I(p^*(\beta),\beta)$ strictly increases in β.*
(ii) If $p^*(\beta) < \overline{p}$, then optimal revenue $\Pi^I(p^*(\beta),\beta)$ strictly decreases in β.
(iii) $p^*(\beta) = \overline{p}$, then optimal revenue $\Pi^I(p^*(\beta),\beta) = \frac{1}{2}\lambda\overline{p}$ is constant in β.

By Proposition 3 and Corollary 1, we know that the optimal revenue $\Pi^I(p^*(\beta),\beta)$ strictly increases in β as β is sufficiently large. Therefore, the revenue-maximizing firm can exploit the bounded rationality when customers are sufficiently irrational.

Finally, we are also interested in how the arrival rate affects the revenue, as it would later be useful. Recall that in Hassin (1986) where customers are fully rational, there exists some λ_0, when $\lambda > \lambda_0$, the revenue function is independent of λ. Interestingly, in our case with boundedly rational customers, we have that higher arrival rate λ always leads to *strictly* higher revenue.

The following Proposition characterizes the effect of arrival rate on the optimal revenue.

PROPOSITION 4. *For any fixed price p and irrationality level $\beta > 0$, the equilibrium joining probability is strictly decreasing and the revenue is strictly increasing in the arrival rate λ.*

The result that higher arrival rates lead to lower equilibrium joining probabilities is not surprising since more congestion forces each customer to lower his joining probability. However, the result that more arrivals always lead to more revenue may appear to be surprising. The key insight is that the marginal revenue increment has to be proportional to the marginal joining probability decrement given the equilibrium condition. This proposition implies that for any price p, clearly not necessarily the optimal price, higher arrival rates lead to higher revenue. In particular, higher arrival rates lead to higher *optimal* revenue. Such finding is in stark contrast to Hassin (1986)'s full-rationality case.

3.3. Social Welfare Maximization

We now turn to study the problem from a social planner's perspective. The social planner is interested in maximizing social welfare. In this section, we study the impact of bounded rationality on the social welfare, both when the price is exogenously given and when the social planner charges the welfare-maximizing price.

For any price p and irrationality level β, the social welfare function is denoted as

$$W^I(\varphi(p,\beta)) = \varphi(p,\beta)\lambda R - \frac{\varphi(p,\beta)\lambda}{\mu - \varphi(p,\beta)\lambda}C. \tag{3}$$

For mathematical convenience, we may drop the dependence over $\varphi(p,\beta)$ and write $W^I(p,\beta)$.

The first term of equation (3) is the average benefit the customers receive from the system, and the second term is the average waiting cost incurred by the customers. Note that the price p affects the social welfare only indirectly through the equilibrium joining probability $\varphi(p,\beta)$.

First, observe that the social welfare $W^I(\varphi(p,\beta))$ is strictly concave in $\varphi(p,\beta)$ (Lemma EC.2 in Appendix C). Combining this fact with the characterization of the equilibrium joining probability $\varphi(p,\beta)$, we can characterize how the social welfare behaves as a function of the irrationality level in Proposition 5.

From Proposition 3, one can obtain that $p^*(0) = R(1-\sqrt{\frac{C}{\mu R}}) = R(1-\sqrt{\frac{1}{v_s}})$. Note that when customers are fully rational, the welfare-maximizing price and the revenue-maximizing price coincide.

PROPOSITION 5. *(i) If $p = \overline{p}$, then social welfare $W^I(\varphi(p,\beta))$ is constant for $\beta \geq 0$.*

(ii) If $[p^*(0) - \overline{p}][p - \overline{p}] \leq 0$ and $p \neq \overline{p}$, then social welfare $W^I(\varphi(p,\beta))$ strictly increases for $\beta \geq 0$.

(iii) If $p \in (\min\{p^*(0), \overline{p}\}, \max\{p^*(0), \overline{p}\}) \bigcup \{p^*(0)\}$ and $p^*(0) \neq \overline{p}$, then social welfare $W^I(\varphi(p,\beta))$ strictly decreases for $\beta \geq 0$.

(iv) If $[p^*(0) - \overline{p}][p - p^*(0)] > 0$, then social welfare $W^I(\varphi(p,\beta))$ strictly increases in $[0, \beta_w(p)]$ and strictly decreases in $(\beta_w(p), \infty)$, where

$$\beta_w(p) = \frac{R - p - \sqrt{\frac{CR}{\mu}}}{\ln \frac{\mu - \sqrt{\frac{C\mu}{R}}}{\lambda - \mu + \sqrt{\frac{C\mu}{R}}}}.$$

This proposition fully characterizes the social welfare as a function of the irrationality level. For the first scenario, the joining probability at price $p = \overline{p}$ is precisely half and it is independent of the irrationality level. Thus the social welfare in (i) is not impacted by the irrationality level. For the second scenario, the probability half lies between the joining probability induced by the welfare-maximizing price when customers are fully rational and the joining probability induced by price p when customers' irrationality level is β, so that increasing the irrationality level makes their "distance" smaller. Thus, the social welfare strictly increases as customers are more irrational. For the third scenario, the joining probability induced by the welfare-maximizing price when customers are fully rational is either too high or too low compared to the joining probability induced by price p when customers' irrationality level is β, so that increasing the irrationality level can only make their "distance" further apart. Therefore, the social welfare strictly decreases in the irrationality level β. For the last scenario, the joining probability induced by the welfare-maximizing price when customers are fully rational can be achieved (in the interior). Hence, as the irrationality level increases from zero, the social welfare is "closer" to the optimal social welfare. In this case, the social welfare function is unimodal in the irrationality level, and the first-order condition yields the irrationality level $\beta_w(p)$.

We illustrate Proposition 5 by a numerical study. In Example 1, we set the parameters $R = 15$, $C = 2$, $\mu = 3$, and $\lambda = 3$. For each given price, we compute the social welfare as a function of the irrationality level and depict it in Figure 3. For this example, one can easily compute $\overline{p} = 13.67$ and $p^*(0) = 11.84$. If $p = \overline{p} = 13.67$, then Proposition 5 (i) applies; If $p = 22$, then Proposition 5 (ii) applies; If $p = 13$ or $p = p^*(0) = 11.84$, then Proposition 4 (iii) applies; If $p = 0$, then Proposition 5 (iv) applies. Figure 3 indeed confirms our theoretical prediction in Proposition 5.

In Example 2, we set the parameters $R = 1$, $C = 2$, $\mu = 3$, and $\lambda = 3$. Again, for each given price, we compute the social welfare as a function of the irrationality level as well as the corresponding

Figure 3 Illustration of Proposition 4: Example 1 ($R = 15, C = 2, \mu = 3, \lambda = 3.$)

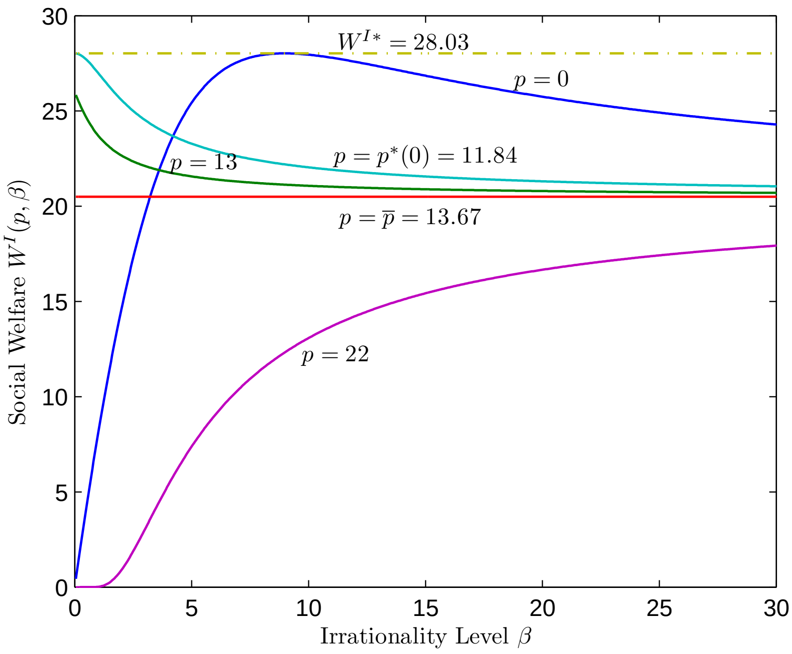
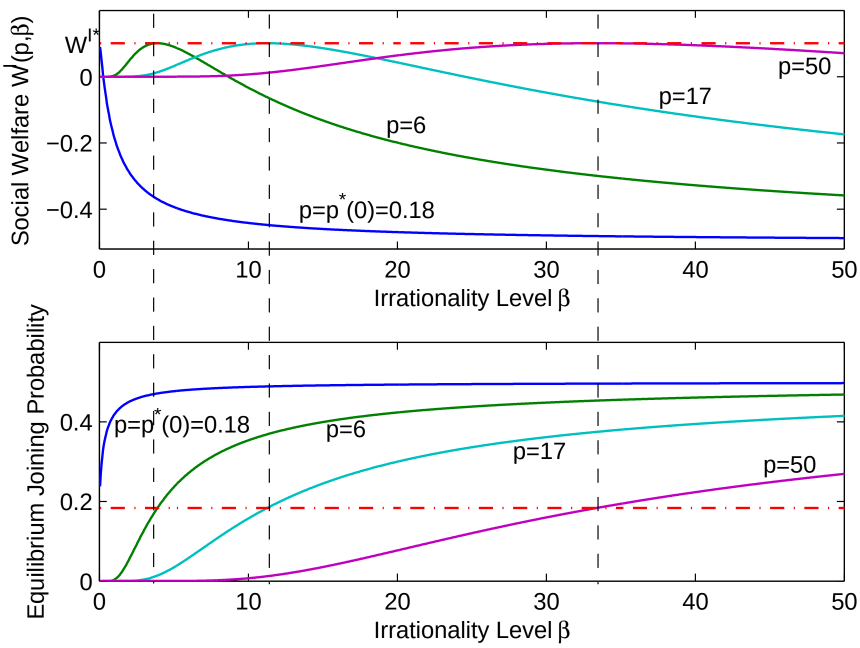


Figure 4 Illustration of Proposition 4: Example 2 ($R = 1, C = 2, \mu = 3, \lambda = 3.$)



equilibrium joining probability and depict them in Figure 4. For this example, one can easily compute $\overline{p} = -0.33$ and $p^*(0) = 0.18$. Observe that for these parameters, case (i) and (ii) of Proposition

5 are not possible, as confirmed in Figure 4.

Proposition 5 implies that the social welfare function is unimodal in the irrational level, as stated in Corollary 2.

COROLLARY 2. *Social welfare $W^I(\varphi(p,\beta))$ is unimodal in the irrationality level β for any price p.*

We are now ready to investigate the optimal social welfare. We prove that the social welfare $W^I(\varphi(p,\beta))$ is unimodal in price p for any irrationality level β (See Lemma EC.4 in Appendix C for a rigorous justification). Finding the welfare-maximizing price boils down to finding the optimal joining probability φ_w^*. To derive the welfare-maximizing price, we use the first-order condition $\frac{\partial W^I(\varphi(p,\beta))}{\partial \varphi(p,\beta)} = 0$ and obtain

$$\varphi_w^* = \frac{\mu - \sqrt{\frac{C\mu}{R}}}{\lambda},$$

which is the optimal equilibrium joining probability that induces the optimal social welfare. Suppose this equilibrium point can be achieved in the interior, then it is required that $R \in (\frac{C}{\mu}, \infty)$ if $\mu < \lambda$; and $R \in (\frac{C}{\mu}, \frac{C\mu}{(\mu-\lambda)^2})$ if $\mu > \lambda$. For cases when the equilibrium point is on the boundary, the problem becomes trivial: if $R \leq \frac{C}{\mu}$, then it is socially optimal to keep everybody out of the system; if $R \geq \frac{C\mu}{(\mu-\lambda)^2}$ when $\mu > \lambda$, then it is socially optimal to let everyone join the system.

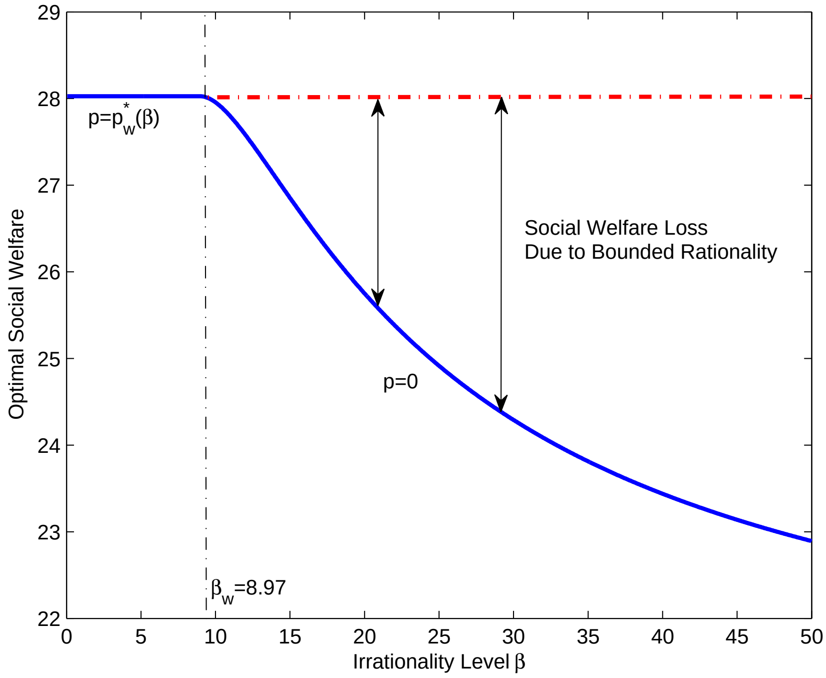
We are now ready to state the welfare-maximizing price that maximizes the social welfare. To state the result, we first substitute the joining probability φ_w^* into the equilibrium condition, i.e., equation (1), and then we obtain the "unconstrained" optimal price by

$$p_w^*(\beta) = R - \sqrt{\frac{CR}{\mu}} - \beta \ln \frac{\mu - \sqrt{\frac{C\mu}{R}}}{\lambda - \mu + \sqrt{\frac{C\mu}{R}}} = p^*(0) - \beta \ln \frac{\varphi_w^*}{1-\varphi_w^*},$$

which can be negative. The welfare-maximizing price is thus $\max\{0, p_w^*(\beta)\}$. Proposition 6 characterizes this welfare-maximizing price and the corresponding social welfare.

PROPOSITION 6. *(i) If $R > \frac{4C\mu}{(2\mu-\lambda)^2}$, when $\beta < \beta_w(0)$ where $\beta_w(0) = \frac{R-\sqrt{\frac{CR}{\mu}}}{\ln \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda-\mu+\sqrt{\frac{C\mu}{R}}}}$, the price $p = p_w^*(\beta)$ is the unique price that maximizes the social welfare, $p_w^*(\beta)$ strictly decreases in β, and the optimal social welfare is $W^I(p_w^*, \beta) = \mu R + C - 2\sqrt{\mu RC}$; when $\beta \geq \beta_w(0)$, the price $p = 0$ is the unique price*

Figure 5 Optimal Social Welfare: Example 1 ($R=15, C=2, \mu=3, \lambda=3, \varphi_w^*=0.7892$)



that yields the maximum social welfare $W^I(0,\beta)$.

(ii) If $R \le \frac{4C\mu}{(2\mu-\lambda)^2}$, *the price* $p=p_w^*(\beta)$ *is the unique price that maximizes the social welfare,* $p_w^*(\beta)$ *strictly increases in* β, *and the optimal social welfare is* $W^I(p_w^*,\beta)=\mu R + C - 2\sqrt{\mu R C}$.

We discuss the implications of this proposition as follows:

If $\varphi_w^* = \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} > \frac{1}{2}$, then $\ln \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda-\mu+\sqrt{\frac{C\mu}{R}}} > 0$, which implies that the price $p_w^*(\beta)$ is strictly decreasing in irrationality level β. In particular, when customers are slightly irrational, the optimal price strictly decreases. The intuition is that the equilibrium joining probability decreases as the irrationality level increases. To achieve the desired optimal joining probability φ_w^*, the social planner has to lower the price as the irrationality level increases.

Similarly, if $\varphi_w^* = \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} < \frac{1}{2}$, then $\ln \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda-\mu+\sqrt{\frac{C\mu}{R}}} < 0$, which implies that the price $p_w^*(\beta)$ is strictly increasing in irrationality level β.

The key insight from this proposition is that the first-best social welfare (which is independent of the irrationality level and the arrival rate) can be achieved when either (i) the optimal joining probability for social welfare maximization is strictly above half and the irrationality level is not too high or (ii) the optimal joining probability for social welfare maximization is below half.

The intuition for this key insight is the following: in these settings, the social planner can always correct for the irrationality on the part of customers, i.e., the social planner still achieves the same optimal social welfare, by charging *appropriate* prices. However, when the optimal joining probability for social welfare maximization is strictly above half and the irrationality level is sufficiently high, the first-best social welfare cannot be achieved. In other words, when the desired joining probability is high, to achieve this, the customers have to join with this probability in equilibrium. However, customers joining probability would be much lower if they are too irrational, and too low even if the firm does not charge any price. In this case, higher irrationality leads to more social welfare losses, as illustrated in Figure 5. This result stands in contrast to the case of revenue maximization.

4. The Visible Queue

4.1. The Model

In this section, we study a single-server queueing system with arrival rate λ and service rate μ with homogenous boundedly rational consumers as before. The only difference from the model setup in Section 3.1 is that the queue length is *visible* to the customers. Hence, every customer does arrive to the system, but he has the option to balk after observing the queue length. Upon arrival, observing n customers in the system, each customer chooses to join the queue with the following logit probability

$$\varphi_n = \frac{e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}}{1+e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}}, \tag{4}$$

for $n = 0, 1, 2, ...$ Observe that the fixed price p always appears as $R - p$ in equation (4). For mathematical convenience, we assume $p = 0$ in this section. However, the findings easily extend to the setting where the price is non-zero. We refer the readers to the Appendix B for a complete treatment.

For ease of exposition, we let $\lambda_n \equiv \lambda\varphi_n$, $n=0,1,2,...$, be the state-dependent arrival rates. Then, we can treat the number of customers in the system as a *birth-death process* with birth rate λ_n and death rate μ. Although customers are boundedly rational, we first show that the *stability* of the system is guaranteed as long as the irrationality level is finite, i.e., $\beta < \infty$, as stated in the following proposition.

PROPOSITION 7. *The visible queueing system with boundedly rational customers is stable for any finite β, and the probability distribution in steady-state is as follows:*

$$P_0 = \frac{1}{1+\Sigma_{k=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k}}$$

is the probability that the system is in state 0, and

$$P_n = \frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu^n(1+\Sigma_{k=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k})}$$

is the probability in state n, $n \geq 1$.

We are interested in how bounded rationality affects social welfare. We can derive the social welfare function as follows:

$$W(\beta) = \Sigma_{n=0}^{\infty}\lambda_n P_n R - \Sigma_{n=0}^{\infty} n P_n C. \tag{5}$$

The first term in equation (5) is the (long run) average reward and the second term is the average waiting cost. Notice that if customers are fully rational, i.e., $\beta = 0$, then our model reduces to Naor (1969)'s model.

In this section, we compare the social welfare $W(\beta)$ and $W(0)$. To compare them, we first define $n_s = [\frac{R\mu}{C}]$ as the threshold queue length used by self-interested customers in deciding to join the queue or not, and n_0, the equivalent threshold from a social planner's point of view. Naor (1969) shows that $n_s \geq n_0$, i.e., self-interested customers typically make the system more congested than the socially optimal level.

Intuitively, irrationality creates two effects for the social welfare, the positive effect and the negative effect. To understand how these two effects come into play, we distinguish the states of the queueing system in three regions: Region 1 refers to the states when the number of customers in the system is less than n_0, Region 2 refers to the states when the number of customers in the system is greater than n_0 but less than n_s, and Region 3 refers to all the other states, i.e., those when the number of customers in the system is greater than n_s. As the customers become more irrational, the joining probabilities of the customers become smaller in Region 1 and 2, but greater in Region 3. The effect from Region 1 is negative, from Region 2 is positive and from Region 3 is negative. As customers become more irrational, these effects take place simultaneously, and it seems unclear *a prior* which effect would dominate.

While equation (5) presents a complete characterization of the social welfare in terms of the irrationality level β, the dependence is quite intricate. Thus we begin by analyzing the social welfare $W(\beta)$ in the neighborhood of zero. We are interested in whether a little bit of irrationality increases or decreases the social welfare, i.e., the relationship between $W(\beta)$ and $W(0)$ when β is sufficiently small. It turns out we are able to completely characterize the conditions in which one effect dominates the other. We have the following simple inequalities showing when the social welfare increases or decreases as the customers become slightly irrational.

PROPOSITION 8. *If any one of the following three conditions is satisfied:*

(1) $n_s < \frac{R\mu}{C} - \frac{1}{2}$*;*

(2) $n_s = n_0$*;*

(3) $n_s = \frac{R\mu}{C} - \frac{1}{2}$ *and* $\rho > 1$*,*

then $W(\beta) < W(0)$ *when* $\beta > 0$ *is sufficiently small. Otherwise,* $W(\beta) > W(0)$ *when* $\beta > 0$ *is sufficiently small.*

According to Proposition 8, if either of the following two conditions is satisfied:

(a) $n_s \neq n_0$, and $n_s > \frac{R\mu}{C} - \frac{1}{2}$,

(b) $n_s \neq n_0$, $n_s = \frac{R\mu}{C} - \frac{1}{2}$ and $\rho \leq 1$,

then a little bit of irrationality **strictly improves** the social welfare.[1]

[1] Finding a simple sufficient and necessary condition for $n_s \neq n_0$ is difficult. However, Lemma EC.5 in Appendix C shows that, either of the two conditions is sufficient for $n_s \neq n_0$: (1) $\rho > 1$ and $n_s > 1$, (2) $\sqrt{2} - 1 < \rho < 1$ and $n_s > 2$.

Notice the logistic probabilities in equation (4), as $\beta \to 0$, the customers who have strictly positive expected utility of joining will join the queue with probability converging to 1, and those who have strictly negative expected utility will join the queue with probability converging to 0. For the sake of a thought experiment, we assume that there is a single state in which customers join with non-degenerate probabilities. If it were the case, it must be the "marginal state" either "on the positive side," i.e., the state where the customer who observes $n_s - 1$ customers in the system, or "on the negative side," i.e., the state where the customer who observes n_s customers in the system. However, in the true system with boundedly rational customers, as long as the customers are slightly irrational, there are multiple states in which customers join the system with non-degenerate probabilities. Considering that the irrationality level is close to zero, it is intuitively clear that it is the *joint effect* of the customer behavior in the marginal state on the positive side and the customer behavior on the negative side that determines the direction of the social welfare change. The effect of the customer behavior in the marginal state on the positive side improves the social welfare, while the effect of the customer behavior in the marginal state on the negative side is detrimental to the social welfare. We have to characterize which effect dominates the other, i.e., to disentangle the joint effect.

The scenario when $n_s = n_0$, i.e., the self-interested (fully rational) customers bring the system to the social optimality, is rare. However, in this setting, a little bit of irrationality causes each customer to randomize and to join the system with non-degenerate probabilities that can only decrease the social welfare. In contrast, when $n_s \neq n_0$, it is the relative location of n_s and $\frac{R\mu}{C} - \frac{1}{2}$ that determines the result. If $n_s > \frac{R\mu}{C} - \frac{1}{2}$, then it is the effect of the customer behavior on the positive side of the marginal state that dominates. Hence, the social welfare is improved. If $n_s < \frac{R\mu}{C} - \frac{1}{2}$, the opposite effect would decrease the social welfare.

The case when $n_s = \frac{R\mu}{C} - \frac{1}{2}$ is more delicate since both effects come into play "simultaneously." It turns out that when $\rho > 1$, the congestion level is so high that the negative effect dominates, and we obtain strictly lower social welfare. When $\rho \leq 1$, the congestion level is low enough to allow the positive effect to dominate, and we obtain strictly higher social welfare.

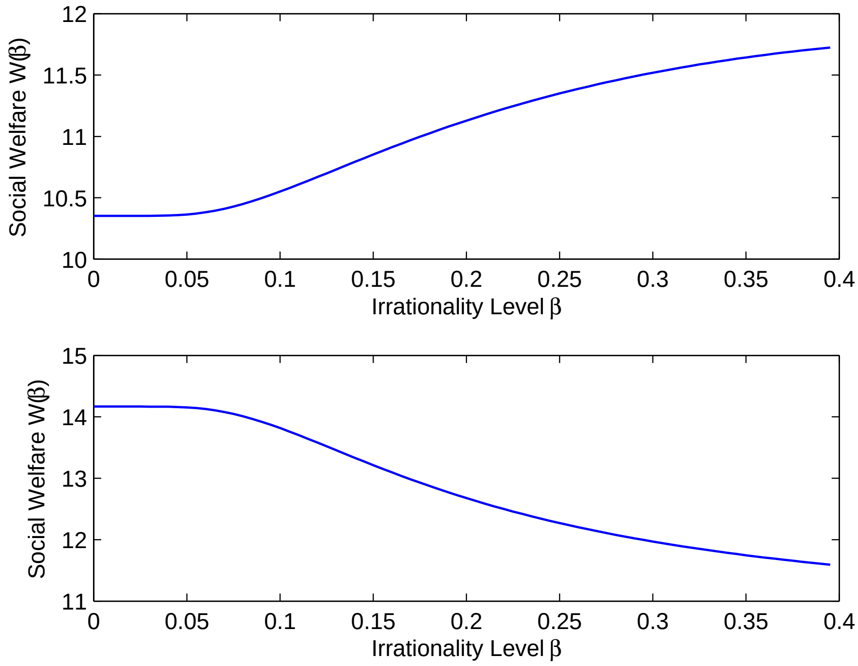
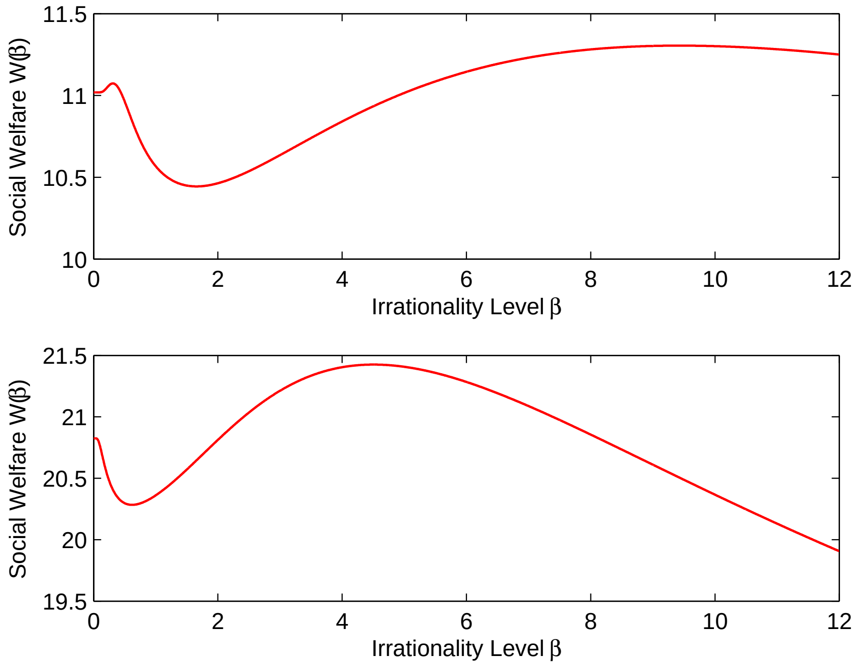


Figure 6 Local Behavior of Social Welfare: Example 3 ($R = 18.96, C = 7, \mu = 3, \lambda = 5, n_s = 8 > \frac{R\mu}{C} - \frac{1}{2} = 7.6257$) and Example 4 ($R = 16, C = 7, \mu = 3, \lambda = 5, v_s = 6.8571, n_s = 6 < \frac{R\mu}{C} - \frac{1}{2} = 6.3571$)

We present a numerical study illustrating this result in Figure 6. We approximate the steady-state distribution by truncating the birth-death process to a finite state space. In this numerical study, we consider two examples. In the first example, we use the following parameters: $R = 18.96, C = 7, \mu = 3, \lambda = 5$. One can easily compute $v_s = 8.1257$, and thus $n_s = 8 > \frac{R\mu}{C} - \frac{1}{2} = 7.6257$. Based on our result in Proposition 8, we expect a little bit of irrationality to improve the social welfare in this case, which is confirmed by the graph in the upper panel in Figure 6. Furthermore, we can see that the social welfare improvement due to irrationality can be non-trivial (more than 10% for example). In the second example, we use the parameters, $R = 16, C = 7, \mu = 3, \lambda = 5$. For this example, $v_s = 6.8571$, and thus $n_s = 6 < \frac{R\mu}{C} - \frac{1}{2} = 6.3571$. For this data, Proposition 8 states that a little bit of irrationality decreases the social welfare, which is in agreement with the graph in the lower panel in Figure 6.

Figure 7 Global Behavior of Social Welfare: Example 5 ($R = 14.93, C = 7, \mu = 3, \lambda = 5, n_s = 6 > \frac{R\mu}{C} - \frac{1}{2} = 5.8986$) and Example 6 ($R = 16, C = 7, \mu = 3, \lambda = 2.6, v_s = 6.8571, n_s = 6 < \frac{R\mu}{C} - \frac{1}{2} = 6.3571$)



In other words, compared to fully rational customers, boundedly rational customers err on both sides, joining a more congested system and balking when congestion is low. While the former is detrimental to the social welfare, the latter can be beneficial. The social welfare can thus be improved depending on which of these effects dominates. In Proposition 8, we provide a simple characterization of this dichotomy. This result appears to be striking: while bounded rationality is usually associated with suboptimal decisions, it might yield better outcomes for the society overall. This is due to the externality present among the boundedly rational customers in the system.

As we discussed before, characterizing the social welfare as a function of the irrationality level is difficult because of the intricate joint effects coming from the three regions simultaneously as customers become more irrational. To understand the social welfare as a function of the irrationality level, we carry out a numerical study. In the first example, the parameters are $R = 14.93, C = 7, \mu = 3, \lambda = 5$, so that $n_s = 6 > \frac{R\mu}{C} - \frac{1}{2} = 5.8986$. As shown in the graph in the upper panel of Figure 7, the social welfare strictly increases initially as predicted by Proposition 8, however, it decreases and then increases again when the customers become more irrational. In the second example, the parameters are $R = 16, C = 7, \mu = 3, \lambda = 2.6$, so that $v_s = 6.8571, n_s = 6 < \frac{R\mu}{C} - \frac{1}{2} = 6.3571$. As illustrated in the graph in the lower panel of Figure 7, the social welfare initially decreases as predicted by Proposition 8, however, it increases as the irrational level becomes larger, and decreases again as the irrationality level further increases. Thus, even though the social welfare is well behaved for a little bit of irrationality, it does not possess global properties such as convexity/concavity or even unimodality. This result stands in contrast to the invisible queue where the social welfare function is unimodal in the irrationality level β.

4.2. Social Welfare Maximization

In this section, we investigate the implication of bounded rationality for the social welfare if the social planner can regulate the system by pricing optimally. We are interested in whether the irrationality increases or decreases the social welfare. We denote the social welfare function $W(p,\beta)$ when the social planner charges price p and customers' irrationality level is β. Obviously, the social welfare $W(p,\beta)$ can be expressed in a similar fashion as equation (5).

Naor (1969) shows that, by "levying tolls" (i.e., charging prices), the social planner can achieve the social optimum when customers are fully rational. In particular, if any price $p^* \in (R - \frac{C(n_0+1)}{\mu}, R - \frac{Cn_0}{\mu}]$ is charged by the social planner, then the maximum social welfare $W^*(0) := \sup_p W(p,0)$ can be achieved. We study whether the optimal social welfare $W^*(0)$ can be achieved by adding irrationality on the part of customers.

We show that when facing boundedly rational customers, the first-best social welfare can never be achieved, as stated in the following proposition.

Proposition 9. *For any price $p \in [0,\infty)$ charged to the customers, the social welfare $W(p,\beta)$ is strictly lower than the social optimum when β is strictly positive, i.e., $W(p,\beta) < W^*(0)$ for $\beta > 0$.*

This proposition proves that irrationality always results in social welfare losses compared to the full-rationality case. This is in contrast to both (a) Naor (1969), where levying tolls achieves the socially optimal welfare and (b) The result in Proposition 8 that a little bit of irrationality can increase the social welfare when an arbitrary price is charged. This stems from the following: When customers are fully rational, the social planner can always regulate the service system by charging prices to achieve the social optimality $W^*(0)$. However, each boundedly rational customer randomizes with non-degenerate probabilities to join or balk. In this case, the social planner loses the *precise* control over the customers' joining decisions, and thus irrationality dilutes the effectiveness of the price regulation.

4.3. Revenue Maximization

In this section, we study the situation in which the firm seeks to maximize its own revenue rather than to optimize the social welfare. We are interested in how bounded rationality impacts the optimal revenue.

The revenue as a function of the price p and the irrationality level $\beta > 0$ is

$$\Pi(p,\beta) = \sum_{n=0}^{\infty} \lambda_n P_n p = \sum_{n=0}^{\infty} \frac{e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}}{1+e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}} \lambda P_n p. \tag{6}$$

When customers are fully rational, i.e., $\beta = 0$, we naturally define $\Pi(p,0) := \lim_{\beta \to 0} \Pi(p,\beta)$, for any price p (one can show such a limit exists). In the setting with fully rational customers, Naor (1969) shows that choosing the revenue-maximizing price boils down to choosing the optimal integer n to maximize the revenue function $\Pi_n = \lambda \frac{1-\rho^n}{1-\rho^{n+1}}(R - \frac{Cn}{\mu})$, so that $p(n) = R - \frac{Cn}{\mu}$. Let n_r be the maximizer, and Π_{n_r} be the maximized revenue.

We are interested in comparing the optimal revenue $\Pi(p^*(\beta),\beta)$ when the revenue-maximizing price $p^*(\beta)$ is set, if customers are slightly irrational, and the optimal revenue Π_{n_r} if customers are fully rational. To compare them, we first partially characterize the optimal price $p^*(\beta)$ when β is sufficiently small. We show that when there is a little bit of irrationality, the firm should charge a strictly *lower* price compared to that under full rationality in order to maximize its revenue (which is made rigorous in Lemma EC.14 in Appendix C). Based on this result, we can show how slight irrationality impacts the optimal revenue as follows.

Proposition 10. *$\Pi(p^*(\beta),\beta) < \Pi_{n_r}$ when β is strictly positive but sufficiently small.*

Proposition 10 implies that a little bit of irrationality *always* results in revenue losses for the firm. However, this result does not extend to situations when the irrationality level becomes high. (Clearly, as customers become fully irrational, the optimal revenue goes to infinity.) The intuition behind Proposition 10 is as follows: A little bit of irrationality forces the revenue-maximizing firm to strictly lower its price compared to full rationality, which in turn brings strictly lower revenue for the firm. In other words, a little bit of irrationality makes revenue-collecting less profitable.

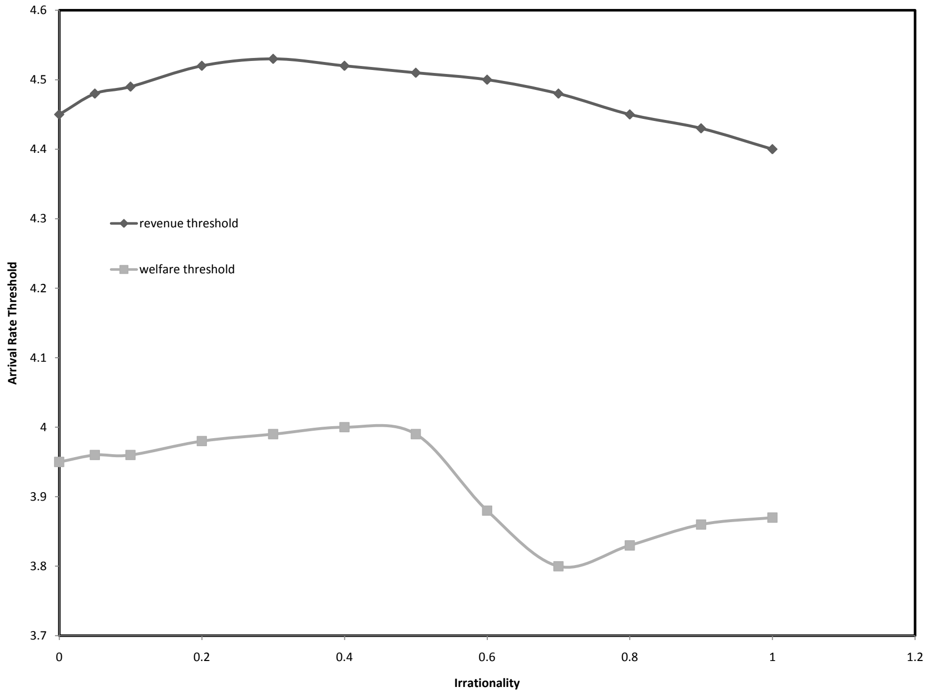
5. Should We Hide the Queue?

In previous sections, we assumed that the queue is either invisible or visible and studied the impact of pricing on social welfare and the revenue of a firm. In this section, we are interested in whether the social planner and the firm should reveal the queue length information if they have the option to do so.

First, it is important to note, if the social planner determines both of the information provision strategy and the pricing decision, it always prefers to reveal the queue length when customers are fully rational (See Naor 1969 and Hassin 1986). When customers are boundedly rational, analytical investigation of this problem is challenging, and thus we turn to numerical studies, which interestingly show the same conclusion.

We next study an analogous problem from a revenue-maximizing firm's point of view, where the firm determines both of the information provision strategy and the pricing decision. Again, we turn to a numerical study. One example is shown in Figure 8, where we have the parameters $R = 60$, $C = 16$ and $\mu = 5$, and we approximate the revenue function by truncating the state space as before. From this numerical study, we have the following observation: For any fixed irrationality level β, there exists a threshold $\overline{\lambda}(\beta)$ such that when $\lambda > \overline{\lambda}(\beta)$, revealing the queue length is preferred by the firm, while $\lambda < \overline{\lambda}(\beta)$, hiding the queue length is preferred. We call the threshold $\overline{\lambda}(\beta)$ the "revenue threshold" which determines when the firm prefers to reveal or hide the queue length for

Figure 8 Arrival Rate Threshold of Hiding the Queue for Social Welfare Maximization ($R = 60, C = 16, \mu = 5$)

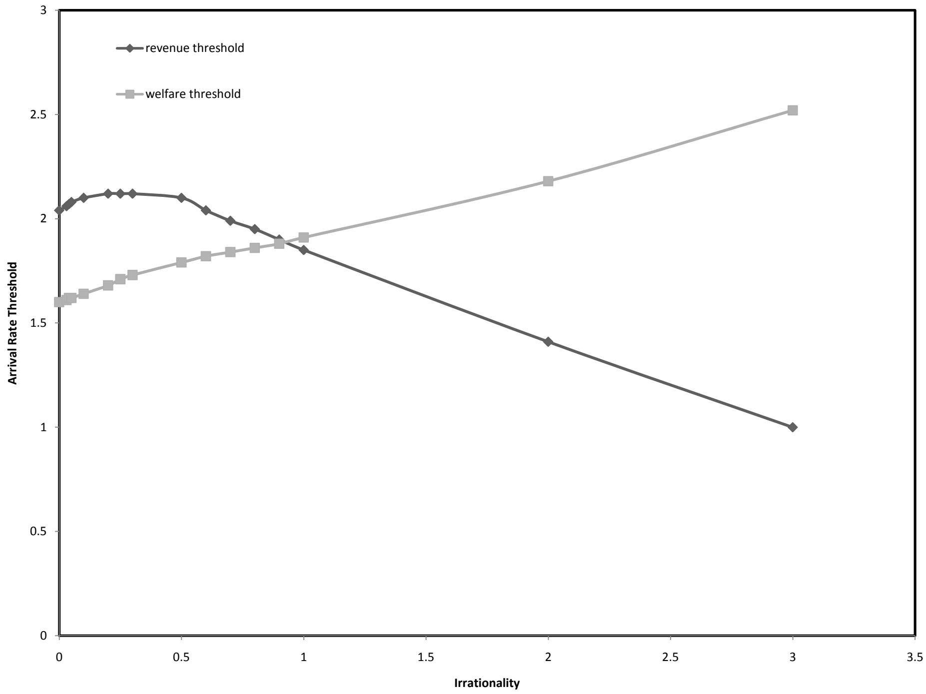


a given irrationality level β. As shown in Figure 8, for these parameters, the revenue threshold $\overline{\lambda}(\beta)$ is not necessarily monotone w.r.t. the irrationality level β.

An intermediate problem arises when the social planner regulates the information provision strategy, yet stops short of regulating prices which allows the revenue maximizer to set those. Assuming customers are fully rational, Hassin (1986) shows that it may be socially optimal to prevent the firm from hiding the queue, but that it is never optimal to encourage the firm to hide the queue length when the firm prefers to reveal the queue length. We are interested in how bounded rationality impacts this conclusion.

Next, we numerically compute the best strategy for the social planner given that the revenue-maximizing firm chooses its optimal prices. From our numerical study, we observe that, there exists a threshold $\widehat{\lambda}(\beta)$ such that, when $\lambda < \widehat{\lambda}(\beta)$, hiding the queue length is preferred; when $\lambda > \widehat{\lambda}(\beta)$, revealing the queue length is preferred. We call $\widehat{\lambda}(\beta)$ the "welfare threshold" which determines when the social planner encourages the firm to reveal or hide the queue. Our numerical results

Figure 9 Arrival Rate Threshold of Hiding the Queue for Social Welfare Maximization ($R=14.93, C=7, \mu=3$)



suggest that $\widehat{\lambda}(\beta)$ again is not necessarily increasing in β, as shown in Figure 8.

Based on the two threshold functions, one can obtain when the social planner and the revenue-maximizing firm align up to reveal or hide the queue. Hassin (1986) studies this problem and show that, in the full-rationality case, it is *never* worthwhile to induce the firm to hide the queue when it does not voluntarily choose it, i.e., the revenue threshold $\overline{\lambda}(0)$ is strictly greater than the welfare threshold $\widehat{\lambda}(0)$. Strikingly, such a conclusion does not generally hold in our setting when customers are boundedly rational. We illustrate this by a numerical example shown in Figure 9, where the parameters are $R=14.93$, $C=7$ and $\mu=3$. For these parameters, when customers become sufficiently irrational, the welfare threshold is higher than the revenue threshold. For instance, when the irrationality level $\beta=2$ and the arrival rate $\lambda=1.7$, the social planner would like to hide the queue, whereas the firm prefers to reveal the queue. Hence, due to bounded rationality, the managerial insight that it is never worthwhile to induce the firm to hide the queue when it does not voluntarily choose it does not extend.

6. Conclusion

The quantal choice paradigm in the behavioral economics literature posits that people are more likely to select better choices than worse ones but do not necessarily succeed in selecting the very best choice. In this paper, we adopt this framework to model bounded rationality and apply it to service systems. Specifically, we study the impact of bounded rationality on the service environment in canonical queueing systems.

For invisible queues, we show the impact of bounded rationality on the revenue, social welfare and the optimal price. For any fixed price, the revenue function is monotone. It can be either decreasing or increasing in the irrationality level depending on whether the equilibrium joining probability is above one half or not. The optimal social welfare is independent of both irrationality level and arrival rate when customers are not "too irrational." When customers are sufficiently irrational, there may be social welfare losses.

For visible queues with a fixed price, we prove that a little bit of irrationality can lead to social welfare improvement, and we provide a simple inequality under which this improvement happens. This result is striking. While bounded rationality is typically associated with suboptimal decisions, the externality among the boundedly rational customers can be beneficial to society. However, as we demonstrated, the relationship between the social welfare (or revenue) and the irrationality level is complex, and fails to be monotone or unimodal. With the optimal prices, we prove that irrationality always decreases social welfare, and a little bit irrationality always results in revenue losses (compared to the full-rationality case). These results tell us that bounded rationality always dilutes the effectiveness of the price regulation in terms of improving the social welfare, and that a little bit of irrationality always makes price-collecting less profitable for the revenue-maximizing firm.

References

Afèche, P. 2004. Incentive-compatible revenue management in queueing systems: optimal strategic delay and other delay tactics. Working paper, University of Toronto.

Allon, G., A. Bassamboo. 2008. Cheap talk in operations: Role of intentional vagueness. *Consumer-Driven Demand and Operations Management Models*, S. Netessine and C. Tang, editors, Springer, Forthcoming.

Armony, M., C. Maglaras. 2004a. Contact center with a call-back option and real-time delay information. *Operations Research* 52 527-545.

Armony, M., C. Maglaras. 2004b. On customer contact centers with a call-back options: Customer decisions, routing rules and system design. *Operations Research* 52 271-292.

Ariely, D. 2008. *Predictably Irrational.* Harper Collins.

Ariely, D. 2009. The end of rational economics. *Harvard Business Review.* July-August.

Arkes, H. R., K. R. Hammond, eds. 1985. *Judgment and Decision Making: An Interdisciplinary Reader.* Cambridge: Cambridge U. Press.

Bajari, P., A. Hortacsu. 2001. Auction models when bidders make small mistakes: Consequences for theory and estimation. *Stanford Economics Working Paper*, CA

Bajari, P., A. Hortacsu. 2005. Are structural estimates of auction models reasonable? Evidence from experimental data, *Journal of Political Economy* 113(4) 703-741.

Basov, S. 2009. Monopolistic screening with boundedly rational consumers. Working paper.

Bendoly, E., R. Croson, P. Goncalves, K. Schultz. 2008. Domains of knowledge useful for behavioral operations management. Working paper, Emory University, Atlanta, GA.

Bendoly, E., K. Donohue, K. Schultz. 2006. Behavioral operations management: Assessing recent findings and revisiting old assumptions. *J. Oper. Management* 24(6) 737-752.

Blume, L. E. 1993. The statistical mechanics of strategic interaction. *Games Econom. Behav.* 5 387-424.

Conlisk, J. 1996. Why bounded rationality? *J. Econom. Literature* 34(2) 669-700.

Fehr, E., K. M. Schmidt. 2006. Chapter 8: The Economics of Fairness, Reciprocity and Altruism-Experimental Evidence and New Theories. *Handbook on the Economics of Giving, Reciprocity and Altruism*, Volume 1, 615-691.

Geigerenzer, G., R. Selten. 2001. *Bounded Rationality: The Adaptive Toolbox.* MIT Press, Cambridge, MA.

Gino, F., G. Pisano. 2008. Toward a theory of behavioral operations. *Manufacturing Service Oper. Management* 10(4) 676-691

Hassin, R. 1986. Consumer information in markets with random products quality: The case of queues and balking. *Econometrica* 54 1185-1195.

Hassin, R. 2009. Equilibrium customers' choice between FCFS and random servers. *Queueing Systems* 62 243-254

Hassin,R., M. Haviv. 2003. *To queue or not to queue: Equilibrium behavior in queueing systems.* Kluwer Academic Publishers

Haviv, M. 2009. Strategic customer behavior in a single server queue. *Wiley Encyclopedia of Operations Research and Management Science.* Forthcoming.

Hogarth, R.. 1980. *Judgment and Choice: Psychology of Decision.* New York, Wiley.

Hsu, V., S. H. Xu, B. Jukic. 2009. Optimal scheduling and incentive compatible pricing for a service system with quality of service guarantees. *Manufacturing Service Oper. Management* 11(3) 375-396.

Kahneman, D., P. Slovic, A. Tversky, eds. 1981. *Judgment under Uncertainty: Heuristic and Biases.* Cambridge, Cambridge U. Press.

Knudsen, N.C. 1972. Individual and social optimization in a multiserver queue with a general cost-benefit structure. *Econometrica* 40 515-528.

Lariviere, M., J. A. Van Mieghem. 2004. Strategically seeking service: How competition can generate Poisson arrivals. *Manufacturing Service Oper. Management* 6(1) 23-40.

Lippman, S. A., S. Stidham Jr. 1977. Individual versus social optimization in exponential congestion systems. *Operations Research* 25(2) 233-247.

Luce, R. D. 1959. *Individual choice behavior: A theoretical analysis.* Wiley, New York.

McKelvey, R. D., T. R. Palfrey. 1995. Quantal response equilibria for normal form games. *Games Econom. Behav.* 10 6-38.

Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* 37 15-24.

Nisbett, R., L. Ross. 1980. *Human inference: Strategies and shortcomings in the social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Payne, J. W., J. R. Bettman, E. J. Johnson. 1992. Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology* 43 87-131.

Pitz, G., N. J. Sachs. 1984. Judgment and decision: Theory and application. *Annual Review of Psychology* 35 139-62.

Plambeck, E. L., Q. Wang. 2010. Implications of hyperbolic discounting for optimal pricing and information management in service systems. Working paper, Stanford University, California.

Rubinstein, A. 1998. *Modeling Bounded Rationality.* MIT Press, Cambrige, MA.

Simon, H. A. 1955. A behavioral model of rational choice. *Quart. J. Econom.* 69(1) 99-118.

Simon, H. 1957. *Models of Man.* John Wiley and Sons, New York.

Su, X. 2008. Bounded rationality in newsvendor models. *Manufacturing Service Oper. Management* 10(4) 566-589.

Tharakunnel, K., S. Bhattacharyya. 2009. Single-leader-multiple-follower games with boundedly rational agents. *Journal of Economic Dynamics & Control* 33 1593-1603.

Thurstone, L. L. 1927. A law of comparative judgment. *Psych. Rev.* 34 273-286.

Tirole, J.. 2007. Bounded rationality and incomplete contracts. Working paper, University of Toulouse and MIT.

Tversky, A., D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185 1124-1131.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *J. Risk Uncertainty* 5(4) 297-323.

Van Mieghem, J. A. 2000. Price and service discrimination in queuing systems: Incentive compatibility of G$c\mu$ scheduling. *Management Sci.* 46(9) 1249-1267.

Yechiali, U. 1971. On optimal balking rules and toll charges in a $GI/M/1$ queuing process. *Operations Research* 19 349-370.

Yechiali, U. 1972. Customers' optimal joining rules for the $GI/M/s$ queue. *Management Sci.* 18 434-443.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Online Appendix for "Bounded Rationality in Service Systems"

This online Appendix has four parts: Appendix A contains the proofs of the propositions in Section 3; Appendix B contains the proofs of the propositions in Section 4. Appendix C contains proofs of all supporting lemmas. Appendix D contains a generalization result of Proposition 8.

EC.1. Appendix A

Proof of Proposition 1. Define $g(\varphi(p,\beta)) := \frac{e^{\frac{R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}}{\beta}}}{1+e^{\frac{R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}}{\beta}}} - \varphi(p,\beta)$, then $g(0) > 0$ and $g(d) = -d < 0$, where $d := \min\{1, \frac{1}{\rho}\}$. $g(\varphi(p,\beta))$ is continuous in $\varphi(p,\beta)$. By the intermediate value theorem, there exists at least one $\varphi^*(p,\beta) \in (0,d)$ such that $g(\varphi^*(p,\beta)) = 0$. Furthermore, $g(\varphi(p,\beta))$ is strictly decreasing in $\varphi(p,\beta)$. Therefore, the solution is unique. ■

Proof of Proposition 2. (i) To prove this part, we need Lemma EC.1 in Appendix C, which states that the equilibrium joining probability is monotone decreasing in the irrationality level if the joining probability is above one half. The reason we need Lemma EC.1 is the following: the fact that $\varphi(p,\beta) > \frac{1}{2}$ is equivalent to $R - p - \frac{C}{\mu-\varphi(p,\beta)\lambda} > 0$, which is equivalent to $R - p - \frac{C}{\mu-\frac{1}{2}\lambda} > 0$, i.e., $p < \overline{p}$. Parts (ii) and (iii) can be shown similarly.

(iv) For fixed β, denote $F(p,\varphi(p)) := \frac{e^{\frac{R-p-\frac{C}{\mu-\varphi(p)\lambda}}{\beta}}}{1+e^{\frac{R-p-\frac{C}{\mu-\varphi(p)\lambda}}{\beta}}} - \varphi(p)$, then equation (1) in the main paper is equivalent to $F(p,\varphi(p)) = 0$. For convenience, we denote $f := f(p,\varphi(p)) = \frac{R-p-\frac{C}{\mu-\varphi(p)\lambda}}{\beta}$. Using the implicit function theorem, we take first derivative w.r.t. p in equation $F(p,\varphi(p)) = 0$. We have

$$\frac{e^f}{\beta(1+e^f)^2} + \frac{e^f \lambda C \varphi'(p)}{\beta(1+e^f)^2(\mu-\varphi(p)\lambda)^2} + \varphi'(p) = 0.$$

Simplify the above, we obtain

$$\varphi'(p) = -\frac{e^f(\mu-\varphi(p)\lambda)^2}{\beta(1+e^f)^2(\mu-\varphi(p)\lambda)^2 + e^f\lambda C} < 0.$$

This completes the proof. ■

Proof of Proposition 3. (i) Recall that $\Pi^I(p,\beta) := p\varphi(p,\beta)\lambda$, where $\varphi(p,\beta)$ is the unique solution to equation (1). For any fixed β, we shall abuse notation by writing $\Pi^I(p)$ and $\varphi(p)$ for brevity.

Taking first derivative, we have $\Pi^{I'}(p) = \lambda[\varphi(p) + p\varphi^{'}(p)]$, where $'$ denotes the derivative. We have shown that $\varphi^{'}(p) < 0$. Let $R_h'(p) = 0$, we have $p = -\frac{\varphi(p)}{\varphi'(p)} > 0$. Now we investigate whether this necessary FOC has multiple solutions. Substituting $\varphi'(p)$, we have

$$p = \frac{\lambda C\varphi(p)}{(\mu - \varphi(p)\lambda)^2} + \frac{\beta}{1-\varphi(p)}.$$

We are interested in whether this equation has a unique solution. For exposition convenience, we denote $g(p) := \frac{\lambda C\varphi(p)}{(\mu-\varphi(p)\lambda)^2} + \frac{\beta}{1-\varphi(p)} - p$, then the question is whether $g(p) = 0$ has a unique solution. We claim that $g(p)$ is strictly decreasing in p. It is clear that the first term in the RHS $\frac{\lambda C\varphi(p)}{(\mu-\varphi(p)\lambda)^2}$ is strictly decreasing in p and so are the second and third terms. Hence $g(p)$ is strictly decreasing in p. Note $g(0) > 0$, and $g(\infty) = -\infty$, so there exists a unique p^* such that $g(p^*) = 0$. Finally, one can verify the second-order condition is satisfied.

(ii) Now we show the second part. We know $p^*(\beta)$ solves the following equation

$$p^*(\beta) = \frac{\lambda C\varphi(p^*(\beta),\beta)}{(\mu - \varphi(p^*(\beta),\beta)\lambda)^2} + \frac{\beta}{1-\varphi(p^*(\beta),\beta)}.$$

Using the implicit function theorem and after simplifying, we have

$$p^{*'}(\beta) = \frac{A\frac{\partial\varphi(p,\beta)}{\partial\beta} + (\mu - \varphi(p,\beta)\lambda)^3(1-\varphi(p,\beta))}{(\mu - \varphi(p,\beta)\lambda)^3(1-\varphi(p,\beta))^2 - A\frac{\partial\varphi(p,\beta)}{\partial p}}.$$

Note that $A > 0$ and $\frac{\partial\varphi(p,\beta)}{\partial p} < 0$, we know the denominator is strictly positive. Hence, $p^{*'}(\beta)$ has the same sign as the numerator. If $\frac{\partial\varphi(p,\beta)}{\partial\beta} \geq 0$, i.e., $p^*(\beta) \geq \overline{p}$ (by Proposition 1), then $p^{*'}(\beta) > 0$. Otherwise, the numerator can be negative depending on the parameters. Hence, $p^{*'}(\beta) > 0$ if $p^*(\beta) \geq \overline{p}$; Otherwise, $p^{*'}(\beta)$ has the same sign as $A\frac{\partial\varphi(p,\beta)}{\partial\beta} + (\mu - \varphi(p,\beta)\lambda)^3(1-\varphi(p,\beta))$, where $A = \lambda C(\mu + \lambda\varphi(p,\beta))(1-\varphi(p,\beta))^2 + \beta(\mu - \varphi(p,\beta)\lambda)^3$.

To further simplify the result in terms of primitives such as β, we want to know whether the price $\overline{p}$ (which leads to equilibrium joining probability 0.5 regardless of the irrationality level including full rationality) can be optimal. Suppose it were, then we have the condition by plugging it into the condition the optimal price has to satisfy,

$$R - \frac{2C}{2\mu-\lambda} = \frac{2\lambda C}{(2\mu-\lambda)^2} + 2\beta.$$

Simplifying yields

$$\beta_0 = \frac{1}{2}R - \frac{2\mu C}{(2\mu-\lambda)^2}.$$

If it is positive, i.e., $R \geq \frac{4\mu C}{(2\mu-\lambda)^2}$, we know $\frac{dp^*(\beta_0)}{d\beta} > 0$, since $\frac{\partial\varphi(\overline{p},\beta_0)}{\partial\beta} = 0$. Then, it is clear that for any $\beta > \beta_0$, we have $\frac{dp^*(\beta)}{d\beta} > 0$. On the other hand, if $R < \frac{4\mu C}{(2\mu-\lambda)^2}$, so that $\beta_0 < 0$, one can easily compute $p^*(0) = R(1-\sqrt{\frac{C}{\mu R}})$. Then if $p^*(0) \geq \overline{p}$, which is equivalent to $R \leq \frac{4C\mu}{(2\mu-\lambda)^2}$, we have $\frac{dp^*(0)}{d\beta} > 0$. This then implies that $\frac{dp^*(\beta)}{d\beta} > 0$ for any $\beta \in [0,\infty)$. ■

Proof of Corollary 1. Using the envelope theorem, we have

$$\frac{d\Pi^I(p^*(\beta),\beta)}{d\beta} = p^*(\beta)\lambda\frac{\partial\varphi(p^*(\beta),\beta)}{\partial\beta}.$$

Then all the results (i), (ii) and (iii) follow directly from Proposition 2 (i)-(iii). ■

Proof of Proposition 4. By definition, $\Pi^{I'}(\lambda) = p[\varphi'(\lambda)\lambda + \varphi(\lambda)]$. To determine its sign, we want to first determine the sign of $\varphi'(\lambda)$. There are at least two ways to do this. First, observing the equilibrium condition, equation (1) in the main paper, suppose $\varphi(\lambda)$ is increasing in λ, then it is easy to see that the LHS is decreasing, while the RHS is increasing, a contradiction. Hence, $\varphi^{*'}(\lambda) < 0$. The other way is to derive this derivative using the implicit function theorem. For convenience, denote $f := f(\lambda) = \frac{R-p-\frac{C}{\mu-\varphi(\lambda)\lambda}}{\beta}$, and $F(\lambda,\varphi(\lambda)) = \frac{e^f}{1+e^f} - \varphi(\lambda)$. The equilibrium condition amounts to $F(\lambda,\varphi(\lambda)) = 0$. Taking first derivative and simplifying, we have

$$\frac{Ce^f}{\beta(1+e^f)^2}\frac{\varphi'(\lambda)\lambda+\varphi(\lambda)}{(\mu-\varphi(\lambda)\lambda)^2} + \varphi'(\lambda) = 0,$$

which clearly implies

$$\frac{d\varphi^*(\lambda)}{d\lambda}\frac{d\Pi^I(\lambda)}{d\lambda} < 0,$$

and $\frac{d\varphi^*(\lambda)}{d\lambda} < 0$. ■

Proof of Proposition 5. (i) According to Lemma EC.3 which gives conditions under which the social welfare is increasing or decreasing in the irrationality level in Appendix C: if $\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} = \frac{1}{2}$ and $p = \overline{p}$, then $W^I(\beta)$ is constant for $\beta \geq 0$. Simplifying the conditions yields result (i).

(ii) According to Lemma EC.3: if $(\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} - \frac{1}{2})(p - R + \frac{2C}{2\mu-\lambda}) > 0$, then $W^I(\beta)$ strictly increases for all $\beta \geq 0$. Combining and simplifying these states in terms of $p^*(0)$ and $\overline{p}$ yields the results.

(iii) According to Lemma EC.3: if either $\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} > \frac{1}{2}$ and $p \in [R-\sqrt{\frac{CR}{\mu}}, R-\frac{2C}{2\mu-\lambda})$, or $\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} < \frac{1}{2}$ and $p \in (R-\frac{2C}{2\mu-\lambda}, R-\sqrt{\frac{CR}{\mu}}]$, then $W^I(\beta)$ strictly decreases for all $\beta \geq 0$. Combining these cases together yields the result in (iii).

(iv) According to Lemma EC.3: if either $\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} > \frac{1}{2}$ and $p < \min\{R-\frac{2C}{2\mu-\lambda}, R-\sqrt{\frac{CR}{\mu}}\}$, or $\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda} < \frac{1}{2}$ and $p > \max\{R-\frac{2C}{2\mu-\lambda}, R-\sqrt{\frac{CR}{\mu}}\}$, then $W^I(\beta)$ strictly increases in $[0,\beta_w)$ then strictly decreases in (β_w,∞). Again, combining and simplifying these states in terms of $p^*(0)$ and $\overline{p}$ yields the results. ■

Proof of Proposition 6. (i) Lemma EC.4 in Appendix C shows that the social welfare function is unimodal in the price, which allows us to invoke the first-order condition to find the optimal price. For any fixed $\lambda > 0$ and irrationality level $\beta > 0$, to achieve social optimality, the equilibrium joining probability $\varphi(p,\beta) = \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda}$, plugging which into the equilibrium condition equation (1) in the main paper, we have

$$\frac{e^{\frac{R-p-\sqrt{\frac{CR}{\mu}}}{\beta}}}{1+e^{\frac{R-p-\sqrt{\frac{CR}{\mu}}}{\beta}}} = \frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda}.$$

Solving this equation, we have the unique solution

$$p_w^*(\beta) = R-\sqrt{\frac{CR}{\mu}} - \beta\log\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda-\mu+\sqrt{\frac{C\mu}{R}}}.$$

We can easily calculate the optimal social welfare at the optimal price $p_w^*(\beta)$ if it is positive (which is satisfied when the conditions stated in part (i) on R and β are satisfied). Otherwise, we have to let price be zero to maximize the social welfare. Part (ii) follows similarly. ■

EC.2. Appendix B

Proof of Proposition 7. Let $\lambda_n \equiv \lambda\varphi_n$, $\mu_n \equiv \mu$, then we can treat the number of customers in the queueing system as a *birth-death process* with birth rate λ_n and death rate μ_n. We have the *balance equations*:

$$\lambda_0 P_0 = \mu_1 P_1,$$

$$(\lambda_n+\mu)P_n = \mu P_{n+1} + \lambda_{n-1}P_{n-1}, n \geq 1.$$

Solving these equations, we have the *limiting probabilities*:

$$P_0 = \frac{1}{1+\Sigma_{k=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k}}, P_n = \frac{\lambda_0\lambda_1...\lambda_{n-1}}{\mu^n(1+\Sigma_{k=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k})}, n \geq 1.$$

The necessary and sufficient condition for the existence of limiting probabilities is:

$$\Sigma_{k=1}^{\infty}\frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k} < \infty.$$

Let $a_k = \frac{\lambda_0\lambda_1...\lambda_{k-1}}{\mu^k}$, using the *ratio test*, we have

$$\frac{a_{k+1}}{a_k} = \frac{\lambda_k}{\mu} = \rho\varphi_k \to 0, k \to \infty.$$

The series converges, hence, the condition is always satisfied for $\beta \in (0, \infty)$. ■

Proof of Proposition 8. To show this proposition, we want to study the social welfare function $W(\beta)$ as β is strictly greater than but arbitrarily close to 0, and compare it with $W(0)$.

We start from the case when only the customers on the two marginal states on the positive and the negative side randomize. Let $\sigma(\beta)$ be the probability of joining for the customer who sees $n_s - 1$ customers in the queue in front of him, and $\delta(\beta)$ be the probability of joining for the customer who observes n_s customers in the queue. We omit β for simplicity. Let $u_0 \equiv U_{n_s-1}$ and $u_1 \equiv U_{n_s}$ be their expected utilities of joining respectively.

If $n_s > \frac{R\mu}{C} - \frac{1}{2}$, we have $u_0 \geq 0, u_1 \leq 0, u_0 + u_1 \leq 0$. By Lemma EC.10 which gives conditions under which "less congestion" implies more welfare holds for any number of customers randomizing using logit probabilities, we need to show

$$(1-\sigma)\rho^{n_s} + \sigma(1-\delta)\rho^{n_s+1} + \sigma\delta\rho^{n_s+2} < \rho^{n_s+1}, \tag{EC.1}$$

when β is small. Some algebra tells us that it suffices to show

$$M(\beta) \equiv \sigma(\delta\rho + 1) = \frac{e^{\frac{u_0}{\beta}}}{1+e^{\frac{u_0}{\beta}}}(\rho\frac{e^{\frac{u_1}{\beta}}}{1+e^{\frac{u_1}{\beta}}} + 1) < 1.$$

We want to know the sign of $M'(\beta)$ when β is small. After lengthy algebra, we have

$$\frac{M'(\beta)(1+e^{\frac{u_0}{\beta}})^2(1+e^{\frac{u_1}{\beta}})^2\beta^2}{e^{\frac{u_0}{\beta}}} = -\rho u_1 e^{\frac{u_0+u_1}{\beta}} - (\rho+1)u_0 e^{\frac{2u_1}{\beta}} - [(\rho+2)u_0 + \rho u_1]e^{\frac{u_1}{\beta}} - u_0 < 0,$$

when $\beta \in (0, \beta_1^*)$, where $M'(\beta_1^*) = 0$.

Observing this inequality, we can see that, if $u_0 + u_1 > 0, u_0 > 0, u_1 < 0$, i.e., $n_s < \frac{R\mu}{C} - \frac{1}{2}$, then $\lim_{\beta \to 0} M'(\beta) > 0$. Hence, the social welfare will decrease in this case. If $u_0 + u_1 = 0, u_0 > 0, u_1 < 0$, i.e., $n_s = \frac{R\mu}{C} - \frac{1}{2}$, then $\lim_{\beta \to 0} M'(\beta) > 0$. If $u_0 = 0, u_1 < 0$, then $M'(\beta)$ has the same sign as $-2\rho u_1 e^{\frac{u_1}{\beta}}$ when β is close to 0, i.e., $\lim_{\beta \to 0} M'(\beta) > 0$.

Next, let us consider the case when the customers in the two marginal states on the positive side and two marginal states on the negative side randomize. Let $\sigma_1(\beta), \sigma(\beta), \delta(\beta), \delta_1(\beta)$ be the probabilities of joining for the customers who observe $n_s - 2, n_s - 1, n_s, n_s + 1$ customers respectively. We will also omit their dependence on β for brevity hereafter. We want to show

$$(1-\sigma_1)\rho^{n_s-1} + \sigma_1(1-\sigma)\rho^{n_s} + \sigma_1\sigma(1-\delta)\rho^{n_s+1} + \sigma_1\sigma\delta(1-\delta_1)\rho^{n_s+2} + \sigma_1\sigma\delta\delta_1\rho^{n_s+3} < \rho^{n_s+1}, \quad \text{(EC.2)}$$

when β is small.

One obvious way to show this inequality is to use the similar technique, differentiation, as for two customer randomizing cases. But it turns out to be much more complicated and untractable. Here is a better way. We already knew when $\beta \in (0, \beta_1^*)$, Inequality (EC.1) holds. Then, it suffices to show

$$(1-\sigma_1) + \sigma_1\rho^2 - \sigma_1\sigma\delta\delta_1\rho^3 + \sigma_1\sigma\delta\delta_1\rho^4 < \rho^2$$

which is equivalent to

$$\frac{\sigma_1\sigma\delta\delta_1}{1-\sigma_1} < \frac{\rho^2 - 1}{\rho^3(\rho - 1)}.$$

This inequality can clearly be satisfied for $\beta \in (0, \beta_2^*)$, where β_2^* makes the inequality above equal. Hence, when $\beta \in (0, \min\{\beta_1^*, \beta_2^*\})$, inequality (EC.2) is satisfied.

Before we generalize our result, let us also consider the case when the customers in the three marginal states on the positive side and three marginal states on the negative side randomize. Let $\sigma_2(\beta), \sigma_1(\beta), \sigma(\beta), \delta(\beta), \delta_1(\beta), \delta_2(\beta)$ be the probabilities of joining for those who see $n_s - 3, n_s - 2, ..., n_s + 2$ customers respectively. We need to show

$$(1-\sigma_2)\rho^{n_s-2} + \sigma_2(1-\sigma_1)\rho^{n_s-1} + \sigma_2\sigma_1\sigma(1-\sigma)\rho^{n_s} + \sigma_2\sigma_1\sigma(1-\delta)\rho^{n_s+1} + \sigma_2\sigma_1\sigma\delta(1-\delta_1)\rho^{n_s+2} +$$

$$\sigma_2\sigma_1\sigma\delta\delta_1(1-\delta_2)\rho^{n_s+3} + \sigma_2\sigma_1\sigma\delta\delta_1\delta_2\rho^{n_s+4} < \rho^{n_s+1}, \quad \text{(EC.3)}$$

when β is small. We already knew, when $\beta \in (0, \min\{\beta_1^*, \beta_2^*\})$, inequality (EC.2) is satisfied. Hence, it suffices to show

$$1-\sigma_2+\sigma_2\rho^3-\sigma_2\sigma_1\sigma\delta\delta_1\delta_2\rho^5+\sigma_2\sigma_1\sigma\delta\delta_1\delta_2\rho^6<\rho^3.$$

which is equivalent to

$$\frac{\sigma_2\sigma_1\sigma\delta\delta_1\delta_2}{1-\sigma_2}<\frac{\rho^3-1}{\rho^5(\rho-1)}.$$

This inequality can clearly be satisfied for $\beta \in (0, \beta_3^*)$, where β_3^* makes the inequality above equal. Hence, when $\beta \in (0, \min\{\beta_1^*, \beta_2^*, \beta_3^*\})$, inequality (EC.3) is satisfied.

Clearly, we can proceed as this until the first arrival customer randomizes, i.e., $2n_s$ customers randomize. For the case when any $2n+2$ customers randomize, $n \le n_s - 1$, we have the inequality to be satisfied

$$\frac{\sigma_n...\sigma_2\sigma_1\sigma\delta\delta_1\delta_2...\delta_n}{1-\sigma_n}<\frac{\rho^{n+1}-1}{\rho^{2n+1}(\rho-1)}. \tag{EC.4}$$

This inequality can clearly be satisfied for $\beta \in (0, \beta_n^*)$, where β_n^* makes the inequality above equal. Hence, when $\beta \in (0, \min\{\beta_1^*, \beta_2^*, \beta_3^*, ..., \beta_n^*\})$, we are done.

Now, we need to consider the cases, when the customers in the $2n$ marginal states randomize, where $n > n_s$. For example, for $n = n_s + 1$, we need to show

$$(1-\sigma_{n_s-1})\rho+\sigma_{n_s-1}(1-\sigma_{n_s-2})\rho^2+\sigma_{n_s-1}\sigma_{n_s-2}\sigma(1-\sigma_{n_s-3})\rho^3+...+\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...(1-\delta_{n_s-1})\rho^{2n_s}$$
$$+\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}(1-\delta_{n_s})\rho^{2n_s+1}+\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}\delta_{n_s}\rho^{2n_s+2}<\rho^{n_s+1} \tag{EC.5}$$

when β is small.

Let $x(\beta) \in (\sigma_1, 1)$ be such that $x(\beta) = y + (1-y)\sigma_1$, where $y \in (0,1)$, then it is easy to verify that inequality (EC.3) can be modified to

$$(1-\sigma_2)\rho^{n_s-2}+\sigma_2(1-\sigma_1)\rho^{n_s-1}+\sigma_2\sigma_1\sigma(1-\sigma)\rho^{n_s}+\sigma_2\sigma_1\sigma(1-\delta)\rho^{n_s+1}+\sigma_2\sigma_1\sigma\delta(1-\delta_1)\rho^{n_s+2}+$$
$$\sigma_2\sigma_1\sigma\delta\delta_1(1-\delta_2)\rho^{n_s+3}+\sigma_2\sigma_1\sigma\delta\delta_1\delta_2\rho^{n_s+4}<x(\beta)\rho^{n_s+1} \tag{EC.6}$$

when $\beta \in (0, \beta_1)$ for some $\beta_1 > 0$. In general, inequality (EC.4) can be modified to

$$\frac{\sigma_n...\sigma_2\sigma_1\sigma\delta\delta_1\delta_2...\delta_n}{1-\sigma_n}<\frac{y\rho^{n+1}-1}{\rho^{2n+1}(\rho-1)}. \tag{EC.7}$$

Then, it suffices to show

$$\delta_{n_s}\rho^{2n_s+1}(\rho-1) < (1-y)\rho^{n_s+1}, \tag{EC.8}$$

which is equivalent to

$$\delta_{n_s} < \frac{1-y}{\rho_s^n(\rho-1)}, \tag{EC.9}$$

which can clearly be satisfied for $\beta \in (0, \beta_y)$ for some $\beta_y > 0$.

When the number of marginal states in which the customers randomize goes to infinity (by Proposition 7, the system is always stable), then we need to show the following summable series is less than ρ^{n_s+1}:

$$\begin{aligned}
&(1-\sigma_{n_s-1})\rho + \sigma_{n_s-1}(1-\sigma_{n_s-2})\rho^2 + \sigma_{n_s-1}\sigma_{n_s-2}\sigma(1-\sigma_{n_s-3})\rho^3 + ... + \sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...(1-\delta_{n_s-1})\rho^{2n_s} \\
&+ [\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}(1-\delta_{n_s})\rho^{2n_s+1} + \sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}\delta_{n_s}(1-\delta_{n_s+1})\rho^{2n_s+2} \\
&+ \sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s}\delta_{n_s+1}(1-\delta_{n_s+2})\rho^{2n_s+3} + ...] < \rho^{n_s+1}.
\end{aligned} \tag{EC.10}$$

We can show that the part in [.] can be made less than $(1-y)(1-\sigma_1)\rho^{n_s+1}$ as $\beta \in (0, \varepsilon^*)$ as follows.

$$\begin{aligned}
&\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}(1-\delta_{n_s})\rho^{2n_s+1} + \sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}\delta_{n_s}(1-\delta_{n_s+1})\rho^{2n_s+2} \\
&+ \sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s}\delta_{n_s+1}(1-\delta_{n_s+2})\rho^{2n_s+3} + ... \\
&= \rho^{2n_s+1}\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}[(1-\delta_{n_s}) + \delta_{n_s}(1-\delta_{n_s+1})\rho + \delta_{n_s}\delta_{n_s+1}(1-\delta_{n_s+2})\rho^2 + ...] \\
&\le \rho^{2n_s+1}(\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1})[1 + \delta_{n_s}\rho + \delta_{n_s}^2\rho^2 + ...] \\
&= \rho^{2n_s+1}(\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1})\frac{1}{1-\delta_{n_s}\rho} < (1-y)(1-\sigma_1)\rho^{n_s+1}
\end{aligned} \tag{EC.11}$$

The last inequality comes from

$$\frac{\sigma_{n_s-1}\sigma_{n_s-2}...\sigma\delta\delta_1...\delta_{n_s-1}}{(1-\delta_{n_s}\rho)(1-\sigma_1)} < \frac{\delta\delta_1...\delta_{n_s-1}}{(1-\delta_{n_s}\rho)(1-\sigma_1)} < \frac{1-y}{\rho^{n_s}} \tag{EC.12}$$

which can easily be satisfied as β is small.

The case when $n_s < \frac{R\mu}{C} - \frac{1}{2}$ or $n_s = \frac{R\mu}{C} - \frac{1}{2}$ can be shown by similar arguments using Lemma EC.8 which states the equivalent results for the case when only the customers on the two marginal states randomize with logit probabilities, Lemma EC.10, and Lemma EC.11 which give conditions under which "more congestion" implies less welfare holds for any number of customers randomizing using logit probabilities. The proofs are omitted for brevity. Hence, we complete the proof. ■

Proof of Proposition 9. Given Lemma EC.12 in Appendix C which shows that for any price charged in the interval $(R-\frac{C(n_0+1)}{\mu}, R-\frac{Cn_0}{\mu}]$, the conclusion holds, we only need to show when p is outside of the interval $(R-\frac{C(n_0+1)}{\mu}, R-\frac{Cn_0}{\mu}]$, the conclusion continues to hold.

We use the same argument as Lemma EC.12. We know that W_{n_0} is the optimal social welfare by Yechiali (1971). However, we cannot rule out the case that $W(p,\beta)=W_{n_0}$ for some p from Yechiali (1971)'s results. To rule out the case, we use Haviv and Puterman (1998), who show that the only average optimal stationary policies are of control limit type, that there are at most two and, if there are two, they occur consecutively. This implies that the only gain optimal randomized stationary policies should randomize over the two control limit states if they exist. The argument is simple: for any randomized policy to be optimal, the deterministic policies it has strictly positive probabilities should yield the same average reward. In our setting with randomization using logit probabilities, their result implies that W_{n_0} is strictly larger than any $W(p,\beta)$ when $\beta>0$ since the logit joining probabilities are in the interval (0,1). We complete the proof. ■

Proof of Proposition 10. By Lemma EC.14 which shows that the optimal price under a little bit of irrationality is strictly lower than the optimal price under full rationality, we obtain the following result: $\Pi(p^*(\beta),\beta)=\Pi(p^*-\epsilon_\beta,\beta)<\Pi(p^*-\epsilon_\beta,0)<\Pi_{n_r}$, for $\beta\in(0,\bar{\beta}_{\epsilon_\beta})$ for some $\bar{\beta}_{\epsilon_\beta}>0$.

■

EC.3. Appendix C

Lemma EC.1. *For any fixed price p, $\varphi(p,\beta)$ is strictly decreasing in β when $\varphi(p,\beta)>\frac{1}{2}$, strictly increasing in β when $\varphi(p,\beta)<\frac{1}{2}$, and constant in β when $\varphi(p,\beta)=\frac{1}{2}$.*

Proof of Lemma EC.1. When $\varphi(p,\beta)>\frac{1}{2}$, we have $R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}>0$. We prove the conclusion by contradiction. Suppose $\Pi^I(p,\beta)$ were increasing in β, then $\varphi(p,\beta)$ and thus the LHS of equation (1) in the main paper increases in β. Note that, $R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}$ decreases in $\varphi(p,\beta)$, and thus $\frac{R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}}{\beta}$ decreases as β increases. Hence the RHS of equation (1) decreases, while the LHS $\varphi(p,\beta)$ increases, which is a contradiction. Therefore, $\varphi(p,\beta)$ is decreasing in β when $\varphi(p,\beta)>\frac{1}{2}$. Similar arguments for other two cases.

Another way to prove it is taking first derivative w.r.t. β directly in equation (1) using the implicit function theorem. In fact, we have

$$[1+\frac{\lambda C e^f}{\beta(1+e^f)^2(\mu-\varphi(p,\beta)\lambda)^2}]\frac{\partial\varphi(p,\beta)}{\partial\beta}=-\frac{fe^f}{\beta(1+e^f)^2},$$

where $f:=f(p,\beta,\varphi(p,\beta))=\frac{R-p-\frac{C}{\mu-\varphi(p,\beta)\lambda}}{\beta}$. Hence, $\frac{\partial\varphi(p,\beta)}{\partial\beta}$ has the same sign as $-f$:

$$sign\frac{\partial\varphi(p,\beta)}{\partial\beta}=-signf.$$

Indeed, after simplifying, we have

$$\frac{\partial\varphi(p,\beta)}{\partial\beta}=\frac{-\ln\frac{\varphi(p,\beta)}{1-\varphi(p,\beta)}(\mu-\varphi(p,\beta)\lambda)^2\varphi(p,\beta)(1-\varphi(p,\beta))}{\beta(\mu-\varphi(p,\beta)\lambda)^2+\lambda C\varphi(p,\beta)(1-\varphi(p,\beta))},$$

where we used the fact that $f=\ln\frac{\varphi(p,\beta)}{1-\varphi(p,\beta)}$. Hence, we complete the proof of the lemma. ■

Lemma EC.2. *$W^I(\varphi(p,\beta))$ is strictly concave in $\varphi(p,\beta)$, i.e., $\frac{\partial^2W^I(\varphi(p,\beta))}{\partial(\varphi(p,\beta))^2}<0$.*

Proof of Lemma EC.2. Taking first-order derivative, we have

$$\frac{\partial W^I(\varphi(p,\beta))}{\partial\varphi(p,\beta)}=\lambda R-\frac{C\lambda\mu}{(\mu-\varphi(p,\beta)\lambda)^2}.$$

Taking second-order derivative, we have

$$\frac{\partial^2W^I(\varphi(p,\beta))}{\partial(\varphi(p,\beta))^2}=-\frac{2\lambda^2\mu C}{(\mu-\varphi(p,\beta)\lambda)^3}<0,$$

which completes the proof. ■

Lemma EC.3. *$\frac{dW^I(\beta)}{d\beta}>0$ if $\varphi(p,\beta)>\max\{\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda},\frac{1}{2}\}$ or $\varphi(p,\beta)<\min\{\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda},\frac{1}{2}\}$; $\frac{dW^I(\beta)}{d\beta}<0$ if $\varphi(p,\beta)<\max\{\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda},\frac{1}{2}\}$ and $\varphi(p,\beta)>\min\{\frac{\mu-\sqrt{\frac{C\mu}{R}}}{\lambda},\frac{1}{2}\}$; $\frac{dW^I(\beta)}{d\beta}=0$ otherwise.*

Proof of Lemma EC.3. Using the chain rule, we have

$$\frac{dW^I(\beta)}{d\beta}=\frac{dW^I(\varphi(\beta))}{d\varphi(\beta)}\frac{\varphi(\beta)}{d\beta}=[\lambda R-\frac{C\lambda\mu}{(\mu-\varphi(p,\beta)\lambda)^2}]\frac{\varphi(\beta)}{d\beta},$$

which has the same sign as $-[\lambda R-\frac{C\lambda\mu}{(\mu-\varphi(p,\beta)\lambda)^2}]f$, where $f:=f(\beta)=\frac{R-p-\frac{C}{\mu-\varphi(\beta)\lambda}}{\beta}$. Then, the lemma follows from determining the sign by discussing all possible cases. ■

LEMMA EC.4. *$W^I(p)$ is unimodal in the price p.*

Proof of Lemma EC.4. By the chain rule, we have $W^{I'}(p) = W^{I'}(\varphi(p))\varphi'(p)$, but $\varphi'(p) < 0$, hence, the FOC $W^{I'}(p) = 0$ is equivalent to $W^{I'}(\varphi(p)) = 0$. By Lemma EC.2 in Appendix C, a unique solution $\varphi(p)$ exists. Again since $\varphi'(p) < 0$, a unique solution p^* exists. ■

LEMMA EC.5. *If $n_s \neq n_0$ and all the customers are fully rational except the one who observes $n_s - 1$ customers in the queue and who randomizes with equal probabilities between "join" and "balk," then the social welfare would be strictly improved. Moreover, either of the two conditions is sufficient for $n_s \neq n_0$: (1) $\rho > 1$ and $n_s > 1$; (2) $\sqrt{2} - 1 < \rho < 1$ and $n_s > 2$.*

Proof of Lemma EC.5. In this case where only the customer who observes $n_s - 1$ customers in the queue is boundedly rational, we can derive an explicit expression for the social welfare function as follows. For the *birth-death process*, we have the birth rates, $\lambda_n = \lambda$, when $n = 0, 1, ..., n_s - 2$; $\lambda_n = \frac{1}{2}\lambda$, when $n = n_s - 1$; $\lambda_n = 0$, when $n \geq n_s$. Then, we have:

$$a_n = \begin{cases} \rho^n \text{ if } n < n_s \\ \frac{1}{2}\rho^n \text{ if } n = n_s \\ 0 \text{ if } n > n_s. \end{cases}$$

$$\begin{aligned} A &\equiv \sum_{n=1}^{\infty} a_n = (a_1 + a_2 + ... + a_{n_s-1}) + a_{n_s} \\ &= (\rho + \rho^2 + ... + \rho^{n_s-1}) + \frac{1}{2}\rho^{n_s} \\ &= \frac{\rho(1-\rho^{n_s-1})}{1-\rho} + \frac{1}{2}\rho^{n_s}. \\ P_0 &= \frac{1}{1+A} \\ &= \frac{2(1-\rho)}{2-\rho^{n_s}-\rho^{n_s+1}}. \\ P_n &= a_n P_0, n \geq 1. \\ F &\equiv \sum_{n=0}^{\infty} \lambda_n P_n \\ &= \sum_{n=0}^{n_s} \lambda_n P_n \end{aligned}$$

$$
\begin{aligned}
&= \lambda P_0[(1+\rho+...+\rho^{n_s-2})+\frac{1}{2}\rho^{n_s-1}] \\
&= \lambda\frac{2-\rho^{n_s-1}-\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}}. \\
G &\equiv \sum_{n=0}^{\infty} nP_n \\
&= \sum_{n=0}^{n_s} nP_n \\
&= (\sum_{n=1}^{n_s} n\rho^n - \frac{1}{2}n_s\rho^{n_s})P_0 \\
&= [\frac{\rho^2(1-\rho^{n_s})}{(1-\rho)^2}+\frac{\rho[1-(n_s+1)\rho^{n_s}]}{1-\rho}-\frac{1}{2}n_s\rho^{n_s}]P_0 \\
&= [\frac{\rho^2(1-\rho^{n_s})}{(1-\rho)^2}+\frac{\rho[1-(n_s+1)\rho^{n_s}]}{1-\rho}-\frac{1}{2}n_s\rho^{n_s}]\frac{2(1-\rho)}{2-\rho^{n_s}-\rho^{n_s+1}} \\
&= \frac{\rho}{1-\rho}-\frac{(n_s+1)\rho^{n_s+1}+n_s\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}}.
\end{aligned}
$$

Then, we have the social welfare in this case

$$W_2 = FR - GC = \lambda\frac{2-\rho^{n_s-1}-\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}}R - [\frac{\rho}{1-\rho}-\frac{(n_s+1)\rho^{n_s+1}+n_s\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}}]C.$$

We want to show $W_2 - W(0) > 0$, which is equivalent to $\lambda RK - CT > 0$, where

$$K \equiv \frac{2-\rho^{n_s-1}-\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}} - \frac{1-\rho^{n_s}}{1-\rho^{n_s+1}},$$

$$T \equiv \frac{(n_s+1)\rho^{n_s+1}}{1-\rho^{n_s+1}} - \frac{(n_s+1)\rho^{n_s+1}+n_s\rho^{n_s}}{2-\rho^{n_s}-\rho^{n_s+1}}.$$

Basic algebra yields

$$K = \frac{-(\rho-1)^2\rho^{n_s-1}}{(\rho^{n_s+1}-1)(\rho^{n_s}+\rho^{n_s+1}-2)} < 0$$

and

$$T = \frac{(n_s+1)\rho^{n_s+1}-n_s\rho^{n_s}-\rho^{2n_s+1}}{(\rho^{n_s+1}-1)(\rho^{n_s}+\rho^{n_s+1}-2)}.$$

Then

$$\frac{T}{K} = \frac{(n_s+1)\rho^{n_s+1}-n_s\rho^{n_s}-\rho^{2n_s+1}}{-(\rho-1)^2\rho^{n_s-1}} = \frac{\rho^{n_s+2}+n_s\rho-(n_s+1)\rho^2}{(\rho-1)^2}. \tag{EC.13}$$

But $\lambda RK - CT > 0 \Leftrightarrow \frac{T}{K} > \frac{\lambda R}{C} = \frac{R\mu}{C}\rho = (n_s+\epsilon)\rho$, where $\epsilon \in [0,1)$. Equivalently, we need to show

$$\frac{\rho^{n_s+1}+n_s-(n_s+1)\rho}{(\rho-1)^2} > \frac{R\mu}{C}. \tag{EC.14}$$

We claim that inequality (EC.14) is equivalent to $n_s \neq n_0$. We show this claim as follows.

If inequality (EC.14) holds, we want to show that $n_s \neq n_0$. By definition, n_0 satisfies the two inequalities (18) and (19) on page 20, Naor (1969), which can be transformed into equivalent inequalities (20) and (21) on page 20. If $n_s = n_0$, then inequality (21) in Naor (1969) contradicts with inequality (EC.14) here. Hence, inequality (EC.14) implies that $n_s \neq n_0$.

On the other hand, if $n_s \neq n_0$, we want to show inequality (EC.14) is true. We know $n_s > n_0$, i.e., $n_s \in \{n_0+1, n_0+2, ...\}$. Naor (1969) shows that $W(0) = P(n)$ as a function of n is "discretely unimodal." Hence, we have $P(n_s-1) > P(n_s)$. Simplification yields

$$\frac{\rho^{n_s+1}+n_s-(n_s+1)\rho}{(\rho-1)^2} > \frac{R\mu}{C},$$

which is precisely inequality (EC.14). Hence, $n_s \neq n_0$ implies inequality (EC.14).

Therefore, we have shown that inequality (EC.14) is equivalent to $n_s \neq n_0$ for $\rho \neq 1$.

For $\rho = 1$, we can compute the social welfare directly, $W(0) = \frac{n_s}{n_s+1}\lambda R - \frac{n_s}{2}C$. Then $n_s \neq n_0$ is equivalent to

$$\frac{n_s-1}{n_s}\lambda R - \frac{n_s-1}{2}C > \frac{n_s}{n_s+1}\lambda R - \frac{n_s}{2}C.$$

Simplifying the above, we obtain

$$\frac{n_s(n_s+1)}{2} > \frac{\lambda R}{C}.$$

Taking limits of the LHS of inequality (EC.14) when $\rho \to 1$, we know

$$\lim_{\rho\to 1}\frac{\rho^{n_s+1}+n_s-(n_s+1)\rho}{(\rho-1)^2} = \frac{n_s(n_s+1)}{2}$$

using the L'Hospital rule. Hence, our results hold when $\rho = 1$.

Furthermore, we are interested in sufficient conditions on primitives to ensure that $n_s \neq n_0$ holds. To show inequality (EC.14), it is *sufficient* to show that $\frac{T}{K} \geq (n_s+1)\rho$, which is equivalent to

$$\frac{\rho^{n_s+2}+n_s\rho-(n_s+1)\rho^2}{(\rho-1)^2} \geq (n_s+1)\rho$$

$\Leftrightarrow$

$$\frac{\rho^{n_s+1}-1}{\rho-1} \geq (n_s+1)\rho$$

when $\rho > 1$. But

$$\frac{\rho^{n_s+1}-1}{\rho-1} = \rho^{n_s} + \rho^{n_s-1} + ... + \rho + 1 > (n_s-1)\rho + (\rho^2+1) > (n_s+1)\rho$$

as long as $n_s > 1$.

Now we prove the case when $\rho < 1$, it is equivalent to show

$$\frac{\rho^{n_s+1} + n_s - (n_s+1)\rho}{(\rho-1)^2} > \frac{R\mu}{C} = n_s + \epsilon.$$

Comparing this inequality with equation (21) on page 20 in Naor 1969, we know that it is necessary that $n_s \neq n_0$. And by the same argument as above, we know it is also sufficient that $n_s \neq n_0$.

Again, to find sufficient conditions on the primitives, we offer the following discussion. It is sufficient to show

$$\frac{\rho^{n_s+1}-1}{\rho-1} = \rho^{n_s} + \rho^{n_s-1} + ... + \rho + 1 \leq (n_s+1)\rho.$$

Clearly, if ρ is too small, for example, $\rho < \frac{1}{n_s+1}$, the inequality above cannot hold. Note that we assumed $n_s \geq 2$. We can use induction to show the inequality above. When $n_s = 2$, we are showing $\rho^2 + \rho + 1 \leq 3\rho$, i.e., $(\rho-1)^2 \leq 0$, which cannot be satisfied if $\rho < 1$. Hence, the social welfare decreases if $n_s = 2$ and $\rho < 1$. Now, we consider the case when $n_s = 3$, then the inequality simplifies to $(\rho-1)[\rho-(\sqrt{2}-1)](\rho+1+\sqrt{2}) \leq 0$. When $\sqrt{2}-1 \leq \rho < 1$, the inequality is satisfied. Then, we can use induction to show this inequality holds for any $n_s \geq 3$, if $\sqrt{2}-1 \leq \rho < 1$. This completes our proof. ■

The next lemma generalizes the above result.

LEMMA EC.6. *If $n_s \neq n_0$, and the customer who observes $n_s - 1$ customers already in the queue has any level of irrationality, i.e., he joins with probability $\sigma \in [0,1)$, then the social welfare would be strictly improved.*

Proof of Lemma EC.6. The proof is similar to that for Lemma EC.5. For this case, we have

$$P_0 = \frac{1-\rho}{1-(1-\sigma)\rho^{n_s} - \sigma\rho^{n_s+1}},$$

$$F = \lambda\frac{1-(1-\sigma)\rho^{n_s-1} - \sigma\rho^{n_s}}{1-(1-\sigma)\rho^{n_s} - \sigma\rho^{n_s+1}},$$

$$G = \frac{\rho}{1-\rho} - \frac{\sigma(n_s+1)\rho^{n_s+1} + (1-\sigma)n_s\rho^{n_s}}{1-(1-\sigma)\rho^{n_s} - \sigma\rho^{n_s+1}}.$$

After some basic but tedious algebra, interestingly and surprisingly, we find $\frac{T}{K}$ in this case is the same as equation (EC.13). Hence, we showed the conclusion using the same argument as Lemma EC.5. ■

It is clear that, if the customers "on the negative side," i.e., the customers whose expected utility is strictly less than zero, have some degree of irrationality of joining the queue with some strictly positive probabilities, then the social welfare will deteriorate. We state this simple result formally.

LEMMA EC.7. *If the customer who observes n_s customers in front of him, joins the queue with some positive probability $\delta \in (0,1]$, then the social welfare will decrease for any $\rho \neq 1$.*

Proof of Lemma EC.7. The argument is intuitively simple: if such a customer joins the queue, his *net effect* on the social welfare is strictly negative, thus making the social welfare worse. This loose argument can be made rigorous using a similar technique as before as follows.

In this case, we have $\lambda_n = \lambda$, for $n = 0, 1, ..., n_s - 1$; $\lambda_{n_s} = \delta\lambda$, where $\delta \in (0,1]$; and $\lambda_n = 0$, for $n \geq n_s + 1$. Then,

$$a_n = \begin{cases} \rho^n \text{ if } n < n_s + 1 \\ \delta\rho^n \text{ if } n = n_s + 1 \\ 0 \text{ if } n > n_s + 1. \end{cases}$$

Long algebra gives us

$$F = \lambda\frac{1-(1-\delta)\rho^{n_s} - \delta\rho^{n_s+1}}{1-(1-\delta)\rho^{n_s+1} - \delta\rho^{n_s+2}},$$

$$G = \frac{\rho}{1-\rho} - \frac{(1-\delta)(n_s+1)\rho^{n_s+1} + \delta(n_s+2)\rho^{n_s+2}}{1-(1-\delta)\rho^{n_s+1} - \delta\rho^{n_s+2}}.$$

Then, we have

$$K = \frac{\delta\rho^{n_s}(\rho-1)^2}{(1-\rho^{n_s+1})(1-(1-\delta)\rho^{n_s+1} - \delta\rho^{n_s+2})} > 0$$

and

$$T = \frac{\delta\rho^{n_s+1}[(n_s+1)-(n_s+2)\rho+\rho^{n_s+2}]}{(1-\rho^{n_s+1})(1-(1-\delta)\rho^{n_s+1}-\delta\rho^{n_s+2})}.$$

First, we consider the case when $\rho > 1$. To show $W_2 < W(0)$, it is sufficient to show $\frac{T}{K} \geq (n_s+1)\rho$. But, $\frac{T}{K} \geq (n_s+1)\rho \Longleftrightarrow (n_s+1)-(n_s+2)\rho+\rho^{n_s+2} \geq (n_s+1)(\rho^2-2\rho+1)$, which is in turn equivalent to $\frac{\rho^{n_s+1}-1}{\rho-1} = \rho^{n_s}+\rho^{n_s-1}+...+\rho+1 \geq n_s+1$, which is true, since $\rho > 1$.

Let us consider the case when $\rho < 1$. To show $W_2 < W(0)$, it is sufficient to show $\frac{\rho^{n_s+1}-1}{\rho-1} = \rho^{n_s}+\rho^{n_s-1}+...+\rho+1 \leq n_s+1$, which is clearly true since $\rho < 1$. We complete the proof. ■

Now, we consider the case where both the customers "on the positive side" and "on the negative side" randomize with logit probabilities. We denote $W_2(\beta)$ as the social welfare when these customers randomize.

Lemma EC.8. *Let both of the customer who observes $n_s - 1$ customers in front of him and the customer who observes n_s customers in front of him join the queue with logit probabilities specified in equation (1) while all other customers are fully rational, and the irrationality level β be sufficiently small. If any one of the following three conditions is satisfied: (1) $n_s < \frac{R\mu}{C} - \frac{1}{2}$; (2) $n_s = n_0$; (3) $n_s = \frac{R\mu}{C} - \frac{1}{2}$ and $\rho > 1$, then $W_2(\beta) < W(0)$. Otherwise, $W_2(\beta) > W(0)$.*

Proof of Lemma EC.8. We sketch the main steps here. We know, $\lambda_n = \lambda$, for $n = 0, 1, ..., n_s - 2$; $\lambda_{n_s-1} = \sigma\lambda$, where $\sigma \in [0, 1)$; $\lambda_{n_s} = \delta\lambda$, where $\delta \in (0, 1]$; and $\lambda_n = 0$, for $n \geq n_s + 1$.Then, we have

$$a_n = \begin{cases} \rho^n \text{ if } n < n_s \\ \sigma\rho^n \text{ if } n = n_s \\ \sigma\delta\rho^n \text{ if } n = n_s+1 \\ 0 \text{ if } n > n_s+1. \end{cases}$$

Long algebra yields

$$F = \lambda\frac{1-(1-\sigma)\rho^{n_s-1}-\sigma(1-\delta)\rho^{n_s}-\sigma\delta\rho^{n_s+1}}{1-(1-\sigma)\rho^{n_s}-\sigma(1-\delta)\rho^{n_s+1}-\sigma\delta\rho^{n_s+2}},$$

$$G = \frac{\rho}{1-\rho} - \frac{(1-\sigma)n_s\rho^{n_s}+\sigma(1-\delta)(n_s+1)\rho^{n_s+1}+\sigma\delta(n_s+2)\rho^{n_s+2}}{1-(1-\sigma)\rho^{n_s}-\sigma(1-\delta)\rho^{n_s+1}-\sigma\delta\rho^{n_s+2}}.$$

Then, we have

$$K \equiv \frac{1-(1-\sigma)\rho^{n_s-1}-\sigma(1-\delta)\rho^{n_s}-\sigma\delta\rho^{n_s+1}}{1-(1-\sigma)\rho^{n_s}-\sigma(1-\delta)\rho^{n_s+1}-\sigma\delta\rho^{n_s+2}} - \frac{1-\rho^{n_s}}{1-\rho^{n_s+1}},$$

$$T \equiv \frac{(n_s+1)\rho^{n_s+1}}{1-\rho^{n_s+1}} - \frac{(1-\sigma)n_s\rho^{n_s} + \sigma(1-\delta)(n_s+1)\rho^{n_s+1} + \sigma\delta(n_s+2)\rho^{n_s+2}}{1-(1-\sigma)\rho^{n_s} - \sigma(1-\delta)\rho^{n_s+1} - \sigma\delta\rho^{n_s+2}}.$$

To determine the sign of $W_2(\beta) - W(0) = \lambda RK - CT$, we first need to know the sign of K. Some algebra tells us, when $\sigma = \sigma^* \equiv \frac{1}{1+\delta\rho}$, $K = 0$; when $\sigma > \sigma^*$, $K > 0$; when $\sigma < \sigma^*$, $K < 0$. Then, we can discuss which cases are possible under different assumptions.

First, we assume that $n_s > \frac{R\mu}{C} - \frac{1}{2}$. Under the assumption that β is sufficiently small, if $n_s > \frac{R\mu}{C} - \frac{1}{2}$, i.e., $U_{n_s-1} - 0 < 0 - U_{n_s}$, where $U_{n_s-1} = R - \frac{n_s C}{\mu}$, and $U_{n_s} = R - \frac{(n_s+1)C}{\mu}$, using the logit probability specification in equation (4), we can rule out the cases when $\sigma \in [\sigma^*, 1]$ by contradiction. Suppose it were possible that $\sigma \geq \sigma^*$, when β is small. After substituting the utility functions and simplifying, we have

$$\rho e^{\frac{U_{n_s-1}+U_{n_s}}{\beta}} \geq 1 + \rho e^{\frac{U_{n_s}}{\beta}}. \tag{EC.15}$$

As β goes to zero, the LHS goes to zero while the RHS goes to 1, which is a contradiction. Then, we have $K < 0$, to show $W_2(\beta) > W(0)$, it is equivalent to show $\frac{T}{K} > \frac{\lambda R}{C}$. After lengthy algebra, we have

$$\frac{T}{K} = \frac{\rho^{n_s+2} + n_s\rho - (n_s+1)\rho^2}{(\rho-1)^2} + \frac{\sigma\delta\rho^2(\rho^{n_s+1}-1)}{(\sigma\delta\rho+\sigma-1)(\rho-1)}. \tag{EC.16}$$

For convenience, denote $g(\beta) = \frac{\sigma\delta\rho^2(\rho^{n_s+1}-1)}{(\sigma\delta\rho+\sigma-1)(\rho-1)}$, where σ and δ are the logit probabilities as functions of the irrationality level β, and we have abused notations by omitting β. Since $\sigma < \sigma^*$ when β is small, we know $\sigma\delta\rho + \sigma - 1 < 0$ when β is small. Now, we claim that

$$\lim_{\beta\to 0} g(\beta) = 0.$$

To show this claim, it is sufficient to show

$$\lim_{\beta\to 0} \frac{\sigma-1}{\sigma\delta} = -\infty.$$

Indeed, we have

$$\frac{\sigma-1}{\sigma\delta} = -\frac{1 + e^{\frac{U_{n_s-1}}{\beta}} + e^{\frac{U_{n_s}}{\beta}} + e^{\frac{U_{n_s-1}+U_{n_s}}{\beta}}}{e^{\frac{U_{n_s-1}+U_{n_s}}{\beta}} + e^{\frac{2U_{n_s-1}+U_{n_s}}{\beta}}}.$$

Multiplying the RHS of this equation by $e^{\frac{-U_{n_s-1}}{\beta}}$ for both of the numerator and the denominator, and taking limit, we know it goes to $-\infty$, since the numerator goes to 1 and the denominator goes to 0 as β goes to 0.

Comparing inequality (EC.14) and (EC.16), we know when β is sufficiently small, inequality (EC.16) holds if inequality (EC.14) holds which is equivalent to $n_s \neq n_0$. Hence, under the assumption that $n_s > \frac{R\mu}{C} - \frac{1}{2}$, if and only if $n_s \neq n_0$, $W_2(\beta) > W(0)$, when β is sufficiently small.

If $n_s < \frac{R\mu}{C} - \frac{1}{2}$, we can similarly show that the social welfare will decrease as follows. We have $U_{n_s-1} + U_{n_s} > 0$, which implies that $\sigma > \sigma^*$, which further implies that $K > 0$. We claim that $W_2(\beta) < W(0)$, which is equivalent to $\frac{T}{K} > \frac{\lambda R}{C}$. Studying equation (EC.16) again, now we have $g(\beta) > 0$ since $\sigma > \sigma^*$, and we know

$$\lim_{\beta \to 0} \frac{\sigma - 1}{\sigma \delta} = 0,$$

which further implies that

$$\lim_{\beta \to 0} g(\beta) = \frac{\rho(\rho^{n_s+1} - 1)}{\rho - 1}.$$

By Lemma EC.5, one can show that this inequality holds as long as $\rho \neq 1$ regardless of anything else.

The last case is when $n_s = \frac{R\mu}{C} - \frac{1}{2}$, which implies that $U_{n_s-1} + U_{n_s} = 0$. Then the inequality $\sigma \geq \sigma^*$ is equivalent to

$$\rho \geq 1 + \rho e^{\frac{U_{n_s}}{\beta}}.$$

If $\rho < 1$, then this inequality cannot hold. Therefore, it has to be the case that $\sigma < \sigma^*$ when β is sufficiently small, which implies that $K < 0$. We claim that $W_2(\beta) > W(0)$, which is equivalent to $\frac{T}{K} > \frac{\lambda R}{C}$, which holds if and only if $n_s \neq n_0$.

If $\rho > 1$, when β is sufficiently small, we have $\sigma > \sigma^*$, which implies that $K > 0$. We claim that $W_2(\beta) < W(0)$, which is equivalent to $\frac{T}{K} > \frac{\lambda R}{C}$, which holds regardless of anything else, by Lemma EC.7. We complete the proof. ■

LEMMA EC.9. *Function* $W_0(x) \equiv \lambda \frac{1-\rho^x}{1-\rho^{x+1}} R - [\frac{\rho}{1-\rho} - \frac{\rho \frac{d\rho^{x+1}}{d\rho}}{1-\rho^{x+1}}]C = \lambda \frac{1-\rho^x}{1-\rho^{x+1}} R - [\frac{\rho}{1-\rho} - \frac{(x+1)\rho^{x+1}}{1-\rho^{x+1}}]C$, $x \in [0, +\infty)$, *is unimodal.*

Proof of Lemma EC.9. Taking the first-order derivative and simplifying, we have

$$W_0'(x) = \frac{1}{(1-\rho^{x+1})^2}[\lambda R(\rho-1)\rho^x \log\rho + C\rho^{x+1}(1-\rho^{x+1}+(x+1)\log\rho)].$$

We want to show that the equation $W'(x)=0$ has at most one solution, which implies that $W_0(x)$ is unimodal. Indeed, $W_0'(x)=0$ is equivalent to

$$\rho^{x+1} = (x+1)\log\rho + \frac{\lambda R(\rho-1)\log\rho}{C\rho} + 1,$$

which clearly has at most one solution. ■

LEMMA EC.10. *Assume $n_s \neq n_0$. Let $\rho^k \equiv f(\rho) = (1-\sigma)\rho^{n_s} + \sigma(1-\delta)\rho^{n_s+1} + \sigma\delta\rho^{n_s+2}$, $V \equiv \rho f'(\rho) = (1-\sigma)n_s\rho^{n_s} + \sigma(1-\delta)(n_s+1)\rho^{n_s+1} + \sigma\delta(n_s+2)\rho^{n_s+2}$. If $k \in (n_s, n_s+1)$, then $W_0(k-1) \equiv \lambda\frac{1-\rho^{k-1}}{1-\rho^k}R - [\frac{\rho}{1-\rho} - \frac{V}{1-\rho^k}]C > W_0(n_s) \equiv \lambda\frac{1-\rho^{n_s}}{1-\rho^{n_s+1}}R - [\frac{\rho}{1-\rho} - \frac{(n_s+1)\rho^{n_s+1}}{1-\rho^{n_s+1}}]C$. In general, the property that "less congestion" implies more welfare holds for any number of customers randomizing using logit probabilities.*

Proof of Lemma EC.10. This lemma follows directly from Lemma EC.9. ■

Similarly, we have the following lemma.

LEMMA EC.11. *Assume $n_s \neq n_0$. Let $\rho^k \equiv f(\rho) = (1-\sigma)\rho^{n_s} + \sigma(1-\delta)\rho^{n_s+1} + \sigma\delta\rho^{n_s+2}$, $V \equiv \rho f'(\rho) = (1-\sigma)n_s\rho^{n_s} + \sigma(1-\delta)(n_s+1)\rho^{n_s+1} + \sigma\delta(n_s+2)\rho^{n_s+2}$. If $k \in (n_s+1, n_s+2)$, then $W_0(k-1) \equiv \lambda\frac{1-\rho^{k-1}}{1-\rho^k}R - [\frac{\rho}{1-\rho} - \frac{V}{1-\rho^k}]C < W_0(n_s) \equiv \lambda\frac{1-\rho^{n_s}}{1-\rho^{n_s+1}}R - [\frac{\rho}{1-\rho} - \frac{(n_s+1)\rho^{n_s+1}}{1-\rho^{n_s+1}}]C$. In general, the property that "more congestion" implies less welfare holds for any number of customers randomizing using logit probabilities.*

Proof of Lemma EC.11. This lemma follows directly from Lemma EC.9. ■

LEMMA EC.12. *If the price $p^* \in (R - \frac{C(n_0+1)}{\mu}, R - \frac{Cn_0}{\mu}]$ is charged to the customers, then the social welfare $W(p^*, \beta)$ is lower than the social optimum, i.e., $W(p^*, \beta) < W^*(0)$ for $\beta > 0$.*

Proof of Lemma EC.12. For convenience, we write W_{n_0} and $W^*(0)$ interchangeably for the first-best social welfare. We discuss two cases when the optimal prices when full rationality is assumed by the social planner are charged. The first case is that $n_0+\epsilon_2$ is the global maxima of the function $W_0(x)$ among the continuous interval $[1,n_s]$ (and n_0 is the global maximum among the discrete candidates $\{1,2,...,n_s\}$ as assumed throughout). Let $\delta(\beta)\in(0,1)$ be the probability the customer who sees n_0 will join the queue (all others are fully rational). To show $W_{n_0}(p^*,\beta)<W_{n_0}$, where $W_{n_0}(p^*,\beta)$ is the social welfare when only the customer who sees n_0 customers in front of him is boundedly rational, using the result of equation (EC.14) in Lemma EC.5 (but using n_0 instead of n_s), we know, it is equivalent to show

$$\frac{T}{K}>\frac{\lambda R}{C}=\rho v_s.$$

But

$$\frac{T}{K}=\frac{\rho[(n_0+1)-(n_0+2)\rho+\rho^{n_0+2}]}{(\rho-1)^2}.$$

Hence, we want to show

$$\frac{(n_0+1)-(n_0+2)\rho+\rho^{n_0+2}}{(\rho-1)^2}>v_s=\frac{R\mu}{C},$$

which is precisely the RHS of inequality (22), page 20, Naor (1969). To show $W(p^*,\beta)<W_{n_0}$, simply note that $W_{n_0}(p^*,\beta)\approx W(p^*,\beta)$ when β is sufficiently small by similar arguments in Proposition 8. Finally, it is clear that, if only the customer who sees n_0-1 customers randomize (all others are fully rational), then the social welfare will decrease.

The second case is that $n_0-\epsilon_1$ is the global maxima among the continuous interval $[1,n_s]$. Let $\sigma(\beta)\in(0,1)$ be the probability the customer who sees n_0-1 will join the queue. Similar to the first case, to show $W(p^*,\beta)<W_{n_0}$, using the results in Lemma EC.8 and EC.11, we know, it is equivalent to show

$$\frac{n_0-(n_0+1)\rho+\rho^{n_0+1}}{(\rho-1)^2}<\frac{R\mu}{C},$$

which is precisely the LHS of inequality (22), page 20, Naor (1969), if $\epsilon_1\neq 0$ (the interesting case).

So far we have shown that for small irrationality levels, the result holds, i.e., $W(p^*,\beta)<W_{n_0}$ for $\beta\in(0,\bar{\beta}_{p^*})$ for some $\bar{\beta}_{p^*}>0$.

Now we show the result when β is large in which case the argument above does not apply. However, we know that W_{n_0} is the optimal social welfare by Yechiali (1971). However, we cannot rule out the case that $W(p,\beta) = W_{n_0}$ for some p from Yechiali (1971)'s results. To rule out the case, we use Haviv and Puterman (1998), who show that the only average optimal stationary policies are of control limit type, that there are at most two and, if there are two, they occur consecutively. This implies that the only gain optimal randomized stationary policies should randomize over the two control limit states if they exist. The argument is simple: For any randomized policy to be optimal, the deterministic policies it has strictly positive probabilities should yield the same average reward. In our setting with randomization using logit probabilities, their result implies that W_{n_0} is strictly larger than any $W(p,\beta)$ when $\beta > 0$ since the logit joining probabilities are in the interval (0,1). We complete the proof. ■

LEMMA EC.13. *For the revenue-maximizing price $p^* = R - \frac{Cn_r}{\mu}$ charged to the customers, the maximum revenue when customers are fully rational cannot be achieved when there is a little bit irrationality among the customers in general, for $\beta \in (0, \bar{\beta})$ for some $\bar{\beta} > 0$.*

Proof of Lemma EC.13. When the optimal price is charged, under full-rationality assumption, the customer who observes $n_r - 1$ customers in front of him will join the queue yet with zero utility. However, when there is a little bit irrationality, he will join with probability 1/2. Such change will make the system less congested compared to the fully-rational case. Recall the optimal revenue under full rationality is

$$\Pi_{n_r} = \lambda \frac{1-\rho^{n_r}}{1-\rho^{n_r+1}}(R - \frac{Cn_r}{\mu}).$$

Note that the function $f(x) := \frac{1-\rho^x}{1-\rho^{x+1}}$ is strictly increasing in x when $\rho \neq 1$, so less congestion implies less revenue. We have $\Pi(p^*, \beta) < \Pi_{n_r}$, for $\beta \in (0, \bar{\beta})$ for some $\bar{\beta} > 0$. ■

LEMMA EC.14. *$p^*(\beta) = p^*(0) - \epsilon_\beta$ for some $\epsilon_\beta > 0$ when β is strictly positive but sufficiently small.*

Proof of Lemma EC.14. We exhaust all candidates to prove this result. First, we show that $p^*(\beta)$ cannot be equal to $p^*(0)$ when β is strictly positive but sufficiently small. By Lemma

EC.13, $\Pi(p^*(0),0)=\lim_{\beta\to 0}\Pi(p^*(0),\beta)<\Pi_{n_r}$. Hence, we have $\Pi(p^*(0),\beta)$ in the neighborhood of $\Pi(p^*(0),0)$ when β is small by continuity. Now if we charge price $p=p^*-\epsilon=R-\frac{Cn_r}{\mu}-\epsilon$ for some small $\epsilon>0$, then under full rationality, the customer who sees n_r-1 customers in front of him will join the queue with ϵ utility. With irrationality level β, his joining probability would be $\varphi_{n_r-1}=\frac{e^{\frac{\epsilon}{\beta}}}{1+e^{\frac{\epsilon}{\beta}}}<1$ but can be sufficiently close to 1, which implies less congestion and thus lower revenue. We have $\Pi(p^*-\epsilon,\beta)<\Pi(p^*-\epsilon,0)<\Pi_{n_r}$, for $\beta\in(0,\bar{\beta}_\epsilon)$ for some $\bar{\beta}_\epsilon>\bar{\beta}>0$. However, we know that $\lim_{\beta\to 0}\Pi(p^*-\epsilon,\beta)=\Pi(p^*-\epsilon,0)$, and $\lim_{\epsilon\to 0}\Pi(p^*-\epsilon,0)=\Pi_{n_r}$. Hence, $\Pi(p^*-\epsilon,\beta)$ can be made arbitrarily close to Π_{n_r} when β is small and ϵ is also small. Hence, we have $\Pi(p^*(0),\beta)<\Pi(p^*-\epsilon,\beta)$, when β is small and ϵ is also small. This shows that $p^*(0)$ cannot be the optimal price when customers are slightly irrational.

Next we show that any price taking this form $p=p^*+\epsilon=R-\frac{Cn_r}{\mu}+\epsilon$ for some fixed small $\epsilon>0$ cannot be the optimal price either. Under full rationality, the customer who observes n_r-1 customers in front of him will not join the queue, and the revenue is strictly higher if the price $p_1=R-\frac{C(n_r+1)}{\mu}$ is charged instead since this modification will still induce the same number of customers to join and the revenue per customer is strictly higher. We have $\Pi(p^*+\epsilon,0)<\Pi_{n_r+1}\le\Pi_{n_r}$. Furthermore, we have $\Pi(p^*+\epsilon,\beta)<\Pi_{n_r+1}$ for $\beta\in(0,\bar{\beta}_{\epsilon 1})$ for some $\bar{\beta}_{\epsilon 1}>0$. Hence, $\Pi(p^*+\epsilon,\beta)<\Pi(p^*-\epsilon_1,\beta)$ when β is small and ϵ_1 is also small.

Other prices "faraway from" $p^*(0)$ clearly cannot be the optimal price when β is small. Therefore, $p^*(\beta)=p^*(0)-\epsilon_\beta$ for some $\epsilon_\beta>0$ when β is small.

Finally, we need to verify the existence of the optimal price $p^*(\beta)$. For any $\beta>0$, we know $\Pi(p,\beta)$ is continuous over the closed interval $[p^*(0)-\frac{C}{\mu},p^*(0)]$. Hence, there exists some $p^*(\beta)\in[p^*(0)-\frac{C}{\mu},p^*(0)]$ to maximize $\Pi(p,\beta)$. We complete the proof. ■

EC.4. Appendix D

In this Appendix, we provide a generalization for Proposition 8. For any fixed price p, typically not optimal, we have a result similar to Proposition 8. To state this result, we define

$$\varphi(p,n)=\frac{e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}}{1+e^{\frac{R-p-\frac{(n+1)C}{\mu}}{\beta}}},$$

a customer's probability of joining the system when price p is charged and there are already n customers in the system ahead of him. Clearly, we have $\varphi(0,n)=\varphi_n$. We define $n(p)=[\frac{(R-p)\mu}{C}]$, then we have $n(0)=n_s$. When customers are fully rational, i.e., $\beta=0$, we have the social welfare function

$$W(p,0)=\lambda\frac{1-\rho^{n(p)}}{1-\rho^{n(p)+1}}R-[\frac{\rho}{1-\rho}-\frac{(n(p)+1)\rho^{n(p)+1}}{1-\rho^{n(p)+1}}]C.$$

We are interested in comparing $W(p,0)$ and $W(p,\beta)$ when β is small. As before, we focus on the interesting case when $n_s\neq n_0$. It is clear that we have to compare $n(p)$ with n_0 noting that $n(p)\leq n_s$.

PROPOSITION EC.1. *If $n(p)>n_0$, we have the following result: If any one of the following two conditions is satisfied: (1) $n(p)<\frac{(R-p)\mu}{C}-\frac{1}{2}$; (2) $n(p)=\frac{(R-p)\mu}{C}-\frac{1}{2}$ and $\rho>1$, then $W(p,\beta)<W(p,0)$ when $\beta>0$ is sufficiently small. Otherwise, $W(p,\beta)>W(p,0)$ when $\beta>0$ is sufficiently small.*

If $n(p)<n_0$, we have the following result: If any one of the following two conditions is satisfied: (1) $n(p)<\frac{(R-p)\mu}{C}-\frac{1}{2}$; (2) $n(p)=\frac{(R-p)\mu}{C}-\frac{1}{2}$ and $\rho>1$, then $W(p,\beta)>W(p,0)$ when $\beta>0$ is sufficiently small. Otherwise, $W(p,\beta)<W(p,0)$ when $\beta>0$ is sufficiently small.

The proof of this result is similar to the proof of Proposition 8, and we omit it for brevity. If $n(p)=n_0$, then p has to be one of the optimal prices $p^*\in(R-\frac{C(n_0+1)}{\mu},R-\frac{Cn_0}{\mu}]$, the analysis is in Section 4.2.

References

Haviv, M., M. L. Puterman. 1998. Bias optimality in controlled queueing systems. *Journal of Applied Probability* 35 136-150.

Yechiali, U. 1971. On optimal balking rules and toll charges in a $GI/M/1$ queuing process. *Operations Research* 19 349-370.