Managing Service Systems with Unknown Quality and Customer Anecdotal Reasoning

Hang Ren

School of Management, University College London, Level 38, 1 Canada Square, London E14 5AA, United Kingdom,
hang.ren.13@ucl.ac.uk

Tingliang Huang

Carroll School of Management, Boston College, Chestnut Hill, Massachusetts 02467, United States, tingliang.huang@bc.edu

Kenan Arifoglu

School of Management, University College London, Level 38, 1 Canada Square, London E14 5AA, United Kingdom,
k.arifoglu@ucl.ac.uk

We consider service systems where customers do not know the distribution of uncertain service quality and cannot estimate it fully rationally. Instead, they form their beliefs by taking the average of several anecdotes, the size of which measures their level of bounded rationality. We characterize the customers' joining behavior and the service provider's pricing, quality control, and information disclosure decisions. Bounded rationality induces customers to form different estimates of the service quality and leads the service provider to use pricing as a market segmentation tool, which is radically different from the full rationality setting. As customers gather more anecdotes, the service provider may first decrease and then increase price, and the revenue is U-shaped. Interestingly, a larger sample size may harm consumer surplus, although it always benefits social welfare. When the service provider also has control over quality, we find that it may reduce both quality and price as customers gather more anecdotes. In addition, a high-quality service provider may not disclose quality information if the sample size is small, while a low-quality service provider may disclose if the sample size is large. Furthermore, as the expected waiting cost increases, information non-disclosure is more attractive, thereby highlighting the importance of incorporating customer bounded rationality in congested settings.

Key words: queueing, bounded rationality, behavioral operations, service systems

1. Introduction

Service systems such as call centers, hospitals, restaurants, night clubs, and amusement parks are prevalent in our everyday lives. In many situations, customers need to choose between waiting for service in a queue or balking based on their inference of the service quality. The existing literature (e.g., Debo et al. 2012, Guo et al. 2014) typically assumes that customers make this inference fully rationally, e.g., following the Bayesian rule. However, this is usually challenging for average customers in real-time decision-making. First, Bayesian updating requires customers to have prior knowledge about the service quality. In practice, they may lack such knowledge because of scarce learning opportunities (e.g., diners patronizing a new restaurant) or because they do not have the

relevant expertise (e.g., car owners incapable of evaluating all possible causes of a car breakdown). Second, the calculation involved in Bayesian updating is demanding. In fact, the psychology and economics literature has long argued that people usually do not follow the Bayesian rule in uncertain situations (e.g., Edwards 1968, Kahneman and Tversky 1979, 1984).

Because of the informational and computational challenges underlying Bayesian updating, customers naturally resort to simplified heuristics when estimating service quality. An example of such a heuristic is to ask acquaintances about their past service experiences (i.e., anecdotes) and then expect a service benefit equal to the sample average. Compared to Bayesian updating, this anecdotal reasoning is non-parametric in nature and thus allows customers to estimate service quality without knowing its distributional parameters. The computation involved in anecdotal reasoning is also usually much simpler than applying the Bayesian rule. Furthermore, the worldwide booming of online social media has greatly facilitated information sharing among customers, which makes anecdotal reasoning especially popular. According to surveys by Nielsen (2013, 2015), customers ranked word-of-mouth recommendations as the most trustworthy source of advertising from 2007 to 2015, with its percentage of trust nearly 20% higher than that of the runner-ups (i.e., consumer opinions posted online and editorial content) in 2015.

Although anecdotal reasoning greatly simplifies customers' inferences of service quality, it may lead them to hold incorrect estimates when service quality is intrinsically uncertain.[1] Specifically, anecdotal-reasoning customers attribute the server's performance inferred from anecdotes entirely to its capability rather than luck. Therefore they overestimate (underestimate) service quality if the performance happens to be good (bad) and thus exhibit a joining behavior different from the full rationality setting. This presents new challenges for the management of service systems: (i) As customers gather more anecdotes, how should the service provider (e.g., a new restaurant with growing popularity) adjust price and service quality? (ii) Should the service provider disclose service quality information (e.g., restaurants inviting expert reviews or giving away free drinks/desserts to customers who post food photos on social media)?

To address these questions, we consider an M/M/1 unobservable queue with uncertain service quality. Customers do not know its expected value and estimate it as the average of several anecdotes. The sample size measures their level of bounded rationality: With more anecdotes, customers estimate service quality more accurately and thus their joining behavior converges to the fully rational benchmark where they know the expected service quality.

---

[1] Uncertainty about service quality is prevalent because of the inherent variability in service providers' performance (e.g., a chef may be unable to guarantee perfect timing and seasoning for every dish) and unknown/unpredictable environmental factors (e.g., a pharmacist may be unable to predict all of a drug's side effects).

Using this anecdotal reasoning framework, we characterize customers' equilibrium joining rate and the service provider's pricing, service quality, and information disclosure decisions. Unlike the fully rational benchmark, the service provider uses pricing as a tool to segment customers with different service quality estimates. In particular, a low-quality service provider prices higher than the fully rational benchmark to target the niche customers who considerably overestimate service quality, and a high-quality service provider prices lower to target the mass customers who do not considerably underestimate service quality. As customers gather more anecdotes, the optimal revenue may first decrease and then increase, while social welfare always increases. Interestingly, consumer surplus may decrease due to intensified congestion. We also incorporate the service provider's quality control decision and find that counter-intuitively, it may target a lower quality level as customers gather more anecdotes (i.e., estimate service quality more accurately).

Apart from pricing, in practice service providers may also have the discretion to inform customers of the mean service quality. We characterize this information disclosure decision and find that high-quality service providers may not disclose information if customers are sufficiently boundedly rational, while low-quality service providers may disclose if customers are rational enough. Moreover, a higher congestion cost makes information non-disclosure more attractive. This underscores the importance of considering customer anecdotal reasoning in congested settings.

The remainder of the paper is organized as follows. We provide a literature review in §2, and present our model and preliminary analysis in §3. In §4, we study the service provider's pricing, quality, and information disclosure decisions. In §5, we extend the analysis by considering welfare-maximizing service systems and customer heterogeneity in the sample size. We present concluding remarks in §6. The proofs of lemmas and propositions are relegated to Online Supplement 1.

2. Related Literature

The traditional queueing economics literature incorporates customers' join-or-balk decision by assuming that they perfectly understand the queueing system (Hassin and Haviv 2003, and references therein). This assumption has been relaxed in several recent studies. For example, the behavioral operations management literature studies service systems where customers do not know the realized service quality and update their beliefs about it based on the queue length (Debo et al. 2012, Debo and Veeraraghavan 2014, Guo et al. 2014) or the expected waiting time (Kremer and Debo 2015). Another branch of literature investigates the operations of diagnostic services where customers rely on an expert service provider to identify the type of service they need (Alizamir et al. 2013, Wang et al. 2010). As with our model, the aforementioned research assumes that customers do not know the realized service quality. However, it focuses on the full rationality setting

where customers infer service quality following the Bayesian rule, whereas we consider boundedly rational customers who estimate service quality based on anecdotes.

Recently, researchers start to incorporate customer bounded rationality in service systems. Huang et al. (2013) capture customers' inability to accurately estimate the expected waiting time by introducing a random error term in their waiting time estimates. They derive the revenue and welfare implications of customer bounded rationality for a monopolistic service provider. Li et al. (2016) extend this analysis by considering market competition and the trade-off between service quality and service rate. Huang and Chen (2015) study a service provider's pricing and service rate decisions when faced with customers who estimate the expected waiting cost based on past experiences and anecdotal reasoning. Cui and Veeraraghavan (2016) consider customers who hold arbitrary beliefs about service rate and study the service provider's decision to disclose the true service rate. Although our paper also focuses on customer bounded rationality in service systems, it is fundamentally different from this literature. First, the bounded rationality in Huang et al. (2013), Li et al. (2016), and Cui and Veeraraghavan (2016) stems from customers' inability to perfectly perceive the service rate or accurately calculate the expected waiting time. In our paper, however, customer bounded rationality stems from their *cognitive* limitation of attributing service performance inferred from anecdotes entirely to capability rather than luck. Second, Huang and Chen (2015) focus on the extreme case where customers estimate the expected waiting time from only one anecdote. In contrast, we study customer anecdotal reasoning regarding service quality based on an *arbitrary* number of anecdotes. This allows us to derive insights into the service provider's information disclosure decision and the evolution of its pricing and quality strategies as customers acquire more anecdotes over time. Third, Cui and Veeraraghavan (2016) assume an exogenous price and service quality, whereas we endogenize both.

Researchers have also incorporated decision-makers' bounded rationality in inventory management. For example, Li et al. (2015) consider competing newsvendors' cognitive limitation in the form of inadequately perceiving demand uncertainty. In contrast, we study a different type of bounded rationality (anecdotal reasoning) in a different operational setting (monopolistic service systems).

Our anecdotal reasoning framework is adapted from the $S(k)$-equilibrium proposed by Osborne and Rubinstein (1998). Unlike the classic Nash equilibrium where each player optimizes based on a belief about other players' behaviors, in the $S(k)$-equilibrium she samples each action k times and chooses the one with the highest expected payoff. This equilibrium concept has been widely applied to model customer bounded rationality in several disciplines, including economics (Spiegler

2006a,b, Szech 2011), marketing (Huang and Yu 2014), and operations management (Huang and Chen 2015).

Our paper is also related to the extensive economics literature on quality information disclosure (for a review, see Dranove and Jin 2010). Grossman (1981) and Milgrom (1981) show that all firms should disclose their quality to fully rational customers because they infer non-disclosing firms as having the lowest quality. This "unraveling" result is overturned if information disclosure is costly (Jovanovic 1982), customers ignore or cannot fully understand the disclosed information (Fishman and Hagerty 2003, Hirshleifer et al. 2004), or customers are unsure about the usefulness of a certain quality attribute (Stivers 2004). Our paper complements this literature by showing that customer bounded rationality alone can also overturn the unraveling result.

There is an emerging service operations literature on information disclosure. Hassin (1986, 2007) examines a service provider's decision to disclose its queue length, service quality, service rate, and unit waiting cost. Guo and Zipkin (2007, 2009) analyze the impact of different levels of waiting time information on the service system's performance and consumer surplus. Guo et al. (2011) study a service provider's decision to disclose delay information to customers who estimate waiting time by the entropy-maximization principle. Note that this literature imposes the full rationality assumption on customers, whereas we focus on boundedly rational customers whose level of rationality (given by the sample size) is influenced by the service provider's information disclosure decision.

3. Model and Preliminaries

We consider an M/M/1 unobservable queue with homogeneous customers arriving according to a Poisson process at rate λ (arrival rate or market potential, hereafter). In what follows, we will refer to the service provider (server, manager, etc.) as "he" and each customer as "she." Both the server and the customers are risk-neutral. Upon arrival, a customer chooses between joining the queue at price p or balking to obtain a constant payoff, which we normalize to zero without loss of generality. Joining customers are served on a first-come first-served basis, and the service time is exponentially distributed with rate μ. Waiting for service costs a customer c per unit of time, and completing the service entails a random benefit $\mathscr{R}$ (service quality or valuation, hereafter). For analytical simplicity, we assume that the benefit is normally distributed, i.e., $\mathscr{R} \sim N(R, \sigma^2)$, where R and σ^2 denote the expected benefit and its variance. The benefit is realized upon completion of the service, and *only* the service provider knows its distribution ex ante.

In contrast, the existing queueing economics literature assumes that customers also know the

service quality distribution (see, e.g., Guo et al. 2014). In our setting, this implies that their joining rate $\lambda_r(p)$ is given by (Hassin and Haviv 2003):

$$\lambda_r(p) = \begin{cases} 0, & \text{if } \mu - \frac{c}{(R-p)^+} < 0, \\ \mu - \frac{c}{R-p}, & \text{if } 0 \leqslant \mu - \frac{c}{(R-p)^+} \leqslant \lambda, \\ \lambda, & \text{if } \mu - \frac{c}{(R-p)^+} > \lambda. \end{cases}$$

We will refer to this case as the fully rational benchmark because customers have perfect knowledge of the expected service quality. In what follows, we will derive the joining rate of anecdotal-reasoning customers and show that it includes the fully rational benchmark as a limiting case.

3.1. Customer Anecdotal Reasoning

Building on the established $S(k)$-equilibrium concept (see, e.g., Huang and Yu 2014, Spiegler 2006a,b), we develop customers' anecdotal reasoning framework as follows. Upon arrival, each customer gathers k service quality anecdotes/samples, which we denote as $\mathscr{R}_i$, $i = 1, \ldots, k$. Each anecdote is an independent draw from the service quality distribution, i.e., $\mathscr{R}_i \sim N(R, \sigma^2)$ for all i. The customer estimates service quality as the sample average of all anecdotes. As a result, her service quality estimate $\mathcal{R}$ is given by

$$\mathcal{R} = \left(\sum_{i=1}^{k} \mathscr{R}_i \right) / k \sim N(R, \sigma^2/k).$$

Note that customers' service quality estimates differ across each other and may not coincide with the mean service quality R (i.e., they are indeed boundedly rational). Both results are due to customer anecdotal reasoning: They attribute the service samples solely to the server's capability instead of luck, so some customers overestimate the expected service quality when the sample average happens to be high, while the others underestimate because the sample average happens to be low. Moreover, the sample size k measures customers' level of bounded rationality since their service quality estimates deviate less from the mean service quality as k increases (i.e., $\mathbf{Var}[\mathcal{R}]$ decreases in k). Intuitively, as customers acquire more anecdotes, the sample average is less influenced by luck and thus reflects the mean service quality more accurately.

REMARK 1. In practice, customers may also acquire anecdotes from online ratings/reviews, which are not independent across customers. We have incorporated this type of anecdote as an extension in Online Supplement 2.4.

3.2. Customer Joining Behavior

Based on the service quality estimate $\mathcal{R}$, customers make the join-or-balk decision to maximize their estimated expected payoff. Therefore, unlike the fully rational benchmark where customers play

mixed strategies, in our setting they use a *pure* threshold strategy: All customers with $\mathcal{R} \geqslant p + W$ join the queue and the rest balk, where W denotes their belief about the expected waiting cost. Since $\mathcal{R} \sim N(R, \sigma^2/k)$, the joining rate $\lambda P(\mathcal{R} > p + W)$ is equal to $\lambda \bar{\Phi}(\sqrt{k}(p + W - R)/\sigma)$, where $\bar{\Phi}$ denotes the complementary cumulative distribution function of the standard normal distribution.

As in the traditional queueing economics literature (Hassin and Haviv 2003), we characterize the equilibrium joining rate $\lambda_a^k(p)$ by the condition that customers' beliefs about the expected waiting time are correct. Therefore, using the PASTA property (Wolff 1982), we have $W = \frac{c}{[\mu - \lambda_a^k(p)]^+}$. Substituting this into the expression for the joining rate, we have

$$\lambda_a^k(p) = \lambda \bar{\Phi} \left(\sqrt{k} \left[p + \frac{c}{\mu - \lambda_a^k(p)} - R \right] / \sigma \right), \tag{1}$$

in which $[\mu - \lambda_a^k(p)]^+$ reduces to $\mu - \lambda_a^k(p)$ because $\lambda_a^k(p) > \mu$ cannot constitute an equilibrium—in this case, the expected waiting time goes to infinity and each joining customer can profitably deviate by balking. The next lemma shows that $\lambda_a^k(p)$ as defined by Equation (1) uniquely exists.

LEMMA 1. *For any $p \geqslant 0$,*

(i) A unique $\lambda_a^k(p) \in (0, \min\{\lambda, \mu\})$ exists and strictly decreases in p.

(ii) The equilibrium joining rate $\lambda_a^k(p)$ strictly decreases in k for $R < p + \frac{c}{(\mu - 0.5\lambda)^+}$, strictly increases in k for $R > p + \frac{c}{(\mu - 0.5\lambda)^+}$, and is invariant in k for $R = p + \frac{c}{(\mu - 0.5\lambda)^+}$.

(iii) $\lim_{k \to +\infty} \lambda_a^k(p) = \lambda_r(p)$.

Lemma 1 shows that the equilibrium joining rate is well-defined and converges to the fully rational benchmark as k goes to infinity. Thus, our anecdotal reasoning framework includes the fully rational benchmark as a limiting case. Intuitively, with more service samples, customers estimate the expected service quality more accurately because the sample average is influenced more by the server's capability than by luck.

Lemma 1 also suggests that bounded rationality leads to fewer customers joining high-quality service systems and more customers joining low-quality service systems. To see the intuition, first recall that the service quality estimates are normally distributed (i.e., Bell-shaped) among customers, who apply the threshold strategy of joining (balking) the queue if their service quality estimate is higher (lower) than the total cost (i.e., the price plus the expected waiting cost). In a low-quality (i.e., $R < p + \frac{c}{(\mu - 0.5\lambda)^+}$) service system, the cost threshold is higher than the expected service quality, so only customers who considerably overestimate the service quality (i.e., the right tail of the Bell curve, dubbed as the *niche* customers) choose to join. In a high-quality (i.e., $R > p + \frac{c}{(\mu - 0.5\lambda)^+}$) service system, the cost threshold is lower than the expected service quality, so

all customers except those who considerably underestimate the service quality (i.e., the Bell curve without the left tail, dubbed as the *mass* customers) choose to join. Since bounded rationality lowers the accuracy of customers' service quality estimates (i.e., both tails of the Bell curves expand), it increases the number of niche customers and decreases the number of mass customers. Therefore, the joining rate for low-quality service systems increases in k and the joining rate for high-quality service systems decreases in k.

4. Analysis

In this section, we study the service provider's revenue maximization problem, as given by:

$$\max_{p \geqslant 0} \Pi(p, k) = p \lambda_a^k(p). \tag{2}$$

In §4.1, we characterize the optimal price and its revenue and welfare implications. Then we incorporate the server's quality control decision and examine the impact of the sample size on the optimal price and quality. In §4.2, we analyze the quality information disclosure decision and how it is influenced by the sample size and service system parameters.

4.1. Pricing and Quality Control

In this section, we analyze the service provider's pricing and quality control decisions and how they are influenced by the sample size k. This is of particular interest because it provides insights regarding a service provider's dynamic decisions in a setting where customers acquire more anecdotes over time.[2] As an example, consider a new restaurant with growing popularity (i.e., k is increasing). Our analysis makes recommendations about how the restaurant owner should change the menu prices and the quality of the food and service accordingly.

We first examine the server's pricing decision, as given by (2). Let $p^*(k)$ and $\Pi^*(k)$ denote the optimal price and revenue for a given sample size k, and let p_r^* and Π_r^* denote the fully rational counterparts. The next proposition characterizes the impact of k on $p^*(k)$ and $\Pi^*(k)$.

PROPOSITION 1. *(i) For any $k \geqslant 1$, a unique $p^*(k) > 0$ exists and $\lim_{k \to +\infty} p^*(k) = p_r^*$.*

(ii) The optimal revenue $\Pi^*(k)$ strictly decreases in k for $R < R_1(k)$, strictly increases in k for $R > R_1(k)$, and is invariant in k for $R = R_1(k)$, where $R_1(k) \equiv \frac{c\mu}{[(\mu - 0.5\lambda)^+]^2} + \frac{\sigma}{2\sqrt{k}\Phi'(0)}$ is the mean service quality at which $\lambda_a^k(p^*(k)) = 0.5\lambda$.

(iii) For a sufficiently low λ, there exists a unique quality threshold $R_2(k) > R_1(k)$, which strictly decreases in k, such that $p^*(k)$ strictly decreases in k for $R < R_2(k)$, strictly increases in k for $R > R_2(k)$, and is invariant in k for $R = R_2(k)$.

---

[2] See Online Supplement 2.1 for a discussion of the dynamic decision process.

Proposition 1(i) shows that the revenue maximization problem is well-defined and the optimal price converges to the fully rational counterpart as the amount of quality information grows. Proposition 1(ii)-(iii) examine the impact of the sample size on the optimal price and revenue, which we illustrate in Figure 1.
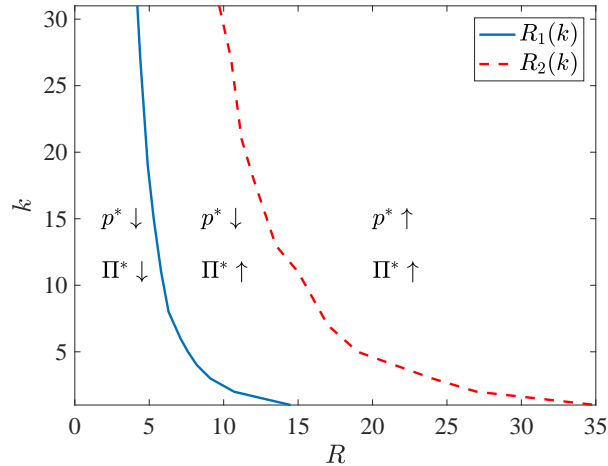

Figure 1   **The Impact of k on p^* and Π^* ($\mu = 2$, $\lambda = 2$, $c = 1$, $\sigma = 1$)**

As shown by Proposition 1(ii) and Figure 1, a larger sample size decreases the optimal revenue of a low-quality (i.e., $R < \underline{R} \equiv \frac{c\mu}{[(\mu-0.5\lambda)^+]^2}$) service provider by reducing demand (i.e., the niche customers), and increases the optimal revenue of a high-quality (i.e., $R > R_1(1)$) service provider by increasing demand (i.e., the mass customers). Somewhat counter-intuitively, an intermediate-quality (i.e., $\underline{R} \leqslant R \leqslant R_1(1)$) service provider's optimal revenue first decreases and then increases in the sample size. This is because the service provider changes the pricing strategy as the number of anecdotes increases. Specifically, as the sample size increases, he switches from targeting exclusively the niche customers to targeting the mass customers. Since a larger sample size decreases demand in the former situation and increases demand in the latter situation, the optimal revenue is U-shaped in k.

The U-shaped revenue is similar in vein to Johnson and Myatt (2006) and Sun (2012). Johnson and Myatt (2006) study the optimal pricing strategy of a firm that sells to customers who hold heterogeneous product valuations. Sun (2012) extends the analysis by considering a 2-period model in which the period-2 customers infer product quality from ratings left by the period-1 customers. Assuming that customers are fully rational and hold heterogeneous product valuations, both papers show that the firm's optimal profit is U-shaped in the variance of customers' service quality estimates. Our paper contributes to this stream of literature by showing that when customers deviate

from full rationality, the U-shaped revenue persists even when customers are ex ante homogeneous in their valuation of service.

Proposition 1(iii) characterizes the impact of the sample size on the service provider's optimal pricing decision. As customers acquire more anecdotes, the number of niche customers decreases while the number of mass customers increases. In response, a low-quality (i.e., $R < R_1(k)$) service provider lowers price and a high-quality (i.e., $R > R_2(k)$) service provider raises price. Surprisingly, we find that an intermediate-quality (i.e., $R_1(k) < R < R_2(k)$) service provider who targets the mass customers lowers price despite the increase in demand. To fully understand this result, note that a larger sample size influences demand in both directions. First, it leads customers to estimate service quality more accurately and thus increases demand. Second, it intensifies congestion and thus decreases demand. When the service quality is not high enough, the second factor dominates and the service provider reduces price to incentivize customers to join. Notably, although we establish Proposition 1(iii) only for the low market potential case, we have numerically verified that it continues to hold when market potential is high (see Figure 1). Moreover, Proposition 1 continues to hold qualitatively for the no-congestion setting since our analysis includes $c = 0$ as a special case.

Managerially, Proposition 1 shows that customer anecdotal reasoning fundamentally changes the optimal pricing strategy in service systems: Managers should use pricing as a market segmentation tool to "milk" the most profitable consumer segment and should change it dynamically as customers acquire more word-of-mouth over time. For example, consider a new restaurant whose service quality is neither too high nor too low. Unlike the widely applied markup-on-cost pricing (Mealey 2016), the owner should start with high prices to target only customers who have received very positive word-of-mouth. As customers gather more anecdotes, he should lower prices to shift from this up-market strategy to the down-market strategy (i.e., even customers receiving occasional negative word-of-mouth patronize). With further growth in the restaurant's popularity, the owner should then switch to raising prices due to the substantial customer base.

4.1.1. Welfare implications. We now investigate the welfare implications of the service provider's optimal pricing strategy. Let $W(p,k)$ and $CS(p,k)$ denote the social welfare and consumer surplus for a given price p and sample size k. By definition, we have

$$W(p,k) = \lambda_a^k(p)R - \frac{c\lambda_a^k(p)}{\mu - \lambda_a^k(p)},$$
$$CS(p,k) = \lambda_a^k(p)R - \frac{c\lambda_a^k(p)}{\mu - \lambda_a^k(p)} - p\lambda_a^k(p),$$

where the first terms on the right-hand side represent the expected social benefit of the service and the second terms are the expected waiting cost. We will denote social welfare by $W(\lambda_a^k(p))$ instead of $W(p,k)$ because it depends on p and k only indirectly through $\lambda_a^k(p)$. This observation suggests that examining the impact of the sample size on social welfare boils down to characterizing its impact on $\lambda_a^k(p^*(k))$, as given by the lemma below.

LEMMA 2. *The optimal joining rate $\lambda_a^k(p^*(k))$ strictly decreases in k for $R < \frac{c\mu}{\{[\mu-\bar{\Phi}(C)\lambda]^+\}^2}$, strictly increases in k for $R > \frac{c\mu}{\{[\mu-\bar{\Phi}(C)\lambda]^+\}^2}$, and is invariant in k for $R = \frac{c\mu}{\{[\mu-\bar{\Phi}(C)\lambda]^+\}^2}$, where $C \approx 0.7517$ is the unique solution of $\bar{\bar{\Phi}}(C) = \Phi'(C)C$.*

Lemma 2 shows that a larger sample size decreases the demand for a low-quality service system (i.e., the niche customers) and increases the demand for a high-quality service system (i.e., the mass customers). Based on this, we examine the impact of k on social welfare and consumer surplus, as shown in the next proposition. For ease of exposition, we will abuse notation and write $W(\lambda_a^k(p^*(k)))$ as $W^*(k)$ and $CS(p^*(k),k)$ as $CS^*(k)$.

PROPOSITION 2. *(i) Social welfare $W^*(k)$ strictly increases in k for $R \neq \frac{c\mu}{\{[\mu-\bar{\Phi}(C)\lambda]^+\}^2}$ and is invariant in k for $R = \frac{c\mu}{\{[\mu-\bar{\Phi}(C)\lambda]^+\}^2}$.*

(ii) Consumer surplus $CS^*(k)$ strictly increases in k for $R < R_1(k)$. If $c > 0$, $CS^*(k)$ strictly decreases in k for a sufficiently high R.

A larger sample size always benefits social welfare because it leads customers to make better decisions by inducing fewer (more) customers to join a low- (high-) quality service system. Somewhat unexpectedly, this improvement in their decision-making at the individual level may harm their aggregate benefit. In fact, Proposition 2(ii) shows that consumer surplus strictly decreases in k when service quality is sufficiently high. The key insight is that the improved decisions (more customers join a high-quality service system) present a negative externality (intensified congestion) for other joining customers. When service quality is high enough, the increased sample size significantly intensifies congestion such that the resulting consumer surplus loss outweighs the consumer surplus benefit due to better decision-making at the individual level. Overall, the consumer surplus decreases. Notably, this result is unique to the congested setting, although the rest of Proposition 2 continues to hold in the absence of congestion (see Online Supplement 2.2). In fact, without congestion a larger sample size never harms consumer surplus since their decision of joining does not impact the other consumers' welfare.

From a managerial perspective, Proposition 2 implies that customers' quality information sharing through word-of-mouth (e.g., dining experiences in a new restaurant) may not create a win-win

situation. In particular, information sharing leads fewer customers to visit a low-quality service system and thus harms the service provider's revenue, whereas it leads more customers to visit a high-quality service system, which intensifies congestion and thus may harm consumer surplus.

4.1.2. Quality control. The preceding analysis focuses on the price as the only decision of the service provider. In practice, he may also have control over the average quality level through, e.g., staff training, facility upgrading, and service design. In this section, we characterize his joint pricing and quality decisions. As before, we focus on the impact of the sample size k on the optimal price and quality because this provides insights regarding a service provider's dynamic quality and pricing strategies in the setting where customers acquire more anecdotes over time (see Online Supplement 2.1 for a detailed discussion).

We incorporate the quality decision by assuming that the service provider can choose R at cost aR^2 (the quality investment, hereafter), where $a > 0$ represents the rate of change of the marginal quality investment. This quadratic cost structure is standard in the product and service design literature (e.g., Anderson et al. 1997, Lahiri and Dey 2013).

The service provider's joint pricing and quality control problem is given by:

$$\max_{p,R \geqslant 0} \ \Pi_R(p,R) = p\lambda_a^k(p,R) - aR^2, \tag{3}$$

where we denote $\lambda_a^k(p)$ as $\lambda_a^k(p,R)$ to stress its dependence on R. The next proposition characterizes the impact of the sample size k on the optimal price $\hat{p}$ and quality $\hat{R}$.

PROPOSITION 3. *(i) A unique $\hat{p} > 0$ determined by $2a\hat{R} = \lambda_a^k(\hat{p}, \hat{R})$ exists.*

(ii) When the market potential λ is sufficiently low, the optimal quality $\hat{R}$ strictly decreases in k. If, in addition, the standard deviation of service quality σ is sufficiently high or sufficiently low, the optimal price $\hat{p}$ strictly decreases in k.

(iii) When the market potential λ is sufficiently high and $c > 0$, the optimal price $\hat{p}$ and quality $\hat{R}$ strictly decrease in k.

(iv) When the market potential λ is sufficiently low, or sufficiently high and $c > 0$, the optimal revenue strictly decreases in k.

A larger sample size implies that customers estimate service quality more accurately. Therefore, conventional wisdom would probably recommend the service provider to improve quality and raise price. However, Proposition 3 points to just the opposite under certain conditions. To fully understand this result, first note that both quality improvement and price reduction may benefit the service provider: Quality improvement allows him to target a higher price, while price reduction

allows him to lower quality and thus save the quality investment. The service provider chooses to reduce price because the quality improvement benefit is lower. Specifically, when market potential is sufficiently low, demand is low and raising price does not increase revenue much. When market potential is sufficiently high, quality improvement leads to significant congestion, so the server cannot raise price much. Notably, this result underscores the importance of incorporating customer bounded rationality in the *congested* setting: Without congestion, the optimal quality and price strictly increase in k under sufficiently high market potential.[3] Intuitively, in this case the server always sets a high quality and targets the mass customers. As they acquire more anecdotes, the joining rate increases and the server improves quality to better exploit the increase in demand, which further allows him to charge a higher price.

We would also like to note that the preceding analysis relies on the implicit assumption that quality control does not impact service quality uncertainty. This holds for certain types of quality improvements (e.g., interior refurbishment and tableware upgrading in a restaurant setting) but may fail for others. If quality improvement reduces service quality uncertainty (e.g., staff training), then our recommendation to reduce both quality and price is strengthened: Quality reduction not only lowers the quality investment, but also increases demand (i.e., the niche customers) by magnifying the variation of customers' service quality estimates. Nevertheless, if quality improvement increases service quality uncertainty (e.g., recipe innovation that requires more complicated cooking techniques), our recommendation may no longer hold.

Proposition 3 characterizes the quality control problem for sufficiently high and low market potential. The intermediate market potential case is analytically challenging, so we resort to a numerical study, as illustrated in Figure 2. Consistent with the exogenous quality scenario, the optimal price and revenue are both U-shaped in k. Perhaps more interestingly, we find that a larger sample size can induce the service provider to improve quality. This is because quality improvement is more profitable than price reduction under intermediate market potential: Quality improvement and the resulting price increase can increase revenue substantially due to the considerable market potential, while it does not significantly intensify congestion because market potential is not too high.

4.2. Quality Information Disclosure

Apart from pricing and quality control, in practice service providers may also have the discretion to disclose service quality information. For example, a hospital can post key performance measures

---

[3] The rest of Proposition 3 continues to hold. See Online Supplement 2.2 for the proof and a related numerical study.
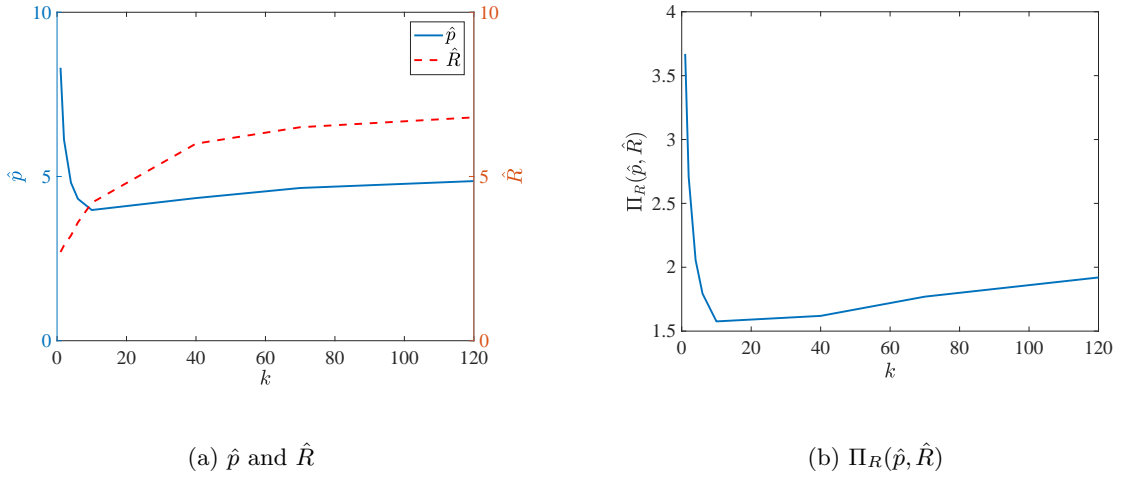
(a) $\hat{p}$ and $\hat{R}$

(b) $\Pi_R(\hat{p}, \hat{R})$

Figure 2 **The Impact of k on $\hat{p}$, $\hat{R}$, and $\Pi_R(\hat{p}, \hat{R})$ under Intermediate Market Potential ($\mu = 2, \lambda = 2, c = 1, \sigma = 10, a = 0.1$)**

(e.g., patient satisfaction scale, readmission rate, and mortality rate) in its waiting rooms, and a restaurant can organize Yelp Elite events (Ayers 2011, Power 2011) and offer free drinks/desserts to customers who post food photos on social media such as Instagram (Dizik 2013), Facebook (Werner 2014), and WeChat (SAMPi 2015).

To incorporate the information disclosure decision, we assume that the service provider chooses to either inform customers of the mean service quality (i.e., $k \to +\infty$) or not (i.e., k is unchanged). Therefore, his joint pricing and information disclosure decisions are given by:

$$\max_{p \geqslant 0, i \in \{k, +\infty\}} \Pi(p, i) = p\lambda_a^i(p), \tag{4}$$

where we ignore the cost of the information disclosure to obtain sharper insights. As in Hassin (2007) and Guo and Zipkin (2007), we focus on truthful information disclosure. This fits our setting since the disclosed information (i.e., the expected service quality) is ex-post verifiable after collecting adequate service samples. Moreover, we do not explicitly consider partial information disclosure (i.e., increasing k to a finite number) because this is always sub-optimal (as will be shown in Proposition 4).

The joint optimization problem in (4) can be treated as a 2-stage optimization problem: The server first chooses k and then chooses p for any given k. By Proposition 1, the optimal price and revenue are $p^*(k)$ and $\Pi^*(k)$. As a result, the server discloses information if $\Pi_r^* > \Pi^*(k)$ and does not disclose if $\Pi_r^* < \Pi^*(k)$. The next proposition provides a full characterization of the information disclosure decision, where $\tilde{k}$ is the sample size determined by $\Pi^*(\tilde{k}) = \Pi_r^*$ and $\tilde{R}$ is the mean service

quality at which $\tilde{k} = 1$.[4]

PROPOSITION 4. *(i) When $R \leqslant \underline{R}$, the service provider does not disclose information.*

(ii)When $\underline{R} < R \leqslant \tilde{R}$, the service provider does not disclose information for $k < \tilde{k}$, discloses information for $k > \tilde{k}$, and is indifferent between the two for $k = \tilde{k}$, where $\tilde{k}$ strictly decreases in R.

(iii) When $R > \tilde{R}$, the service provider discloses information.

We illustrate Proposition 4 by a numerical example in Figure 3. Intuitively, a low-quality (i.e., $R \leqslant \underline{R}$) service provider exploits customer bounded rationality by pricing high to target exclusively the niche customers. Therefore, he chooses not to inform them of the mean service quality. In contrast, a high-quality (i.e., $R > \tilde{R}$) service provider targets the mass customers. Since bounded rationality lowers revenue by leading some customers to balk due to occasional unfavorable anecdotes, the service provider chooses to inform customers of the mean service quality. Interestingly, we find that an intermediate-quality (i.e., $\underline{R} < R \leqslant \tilde{R}$) service provider informs customers of the mean service quality only when they are sufficiently rational (i.e., $k > \tilde{k}$). To see the intuition, first recall from Proposition 1 that the service provider switches from targeting exclusively the niche customers to targeting the mass customers as the sample size increases. Since information disclosure induces all customers to form an accurate service quality estimate (i.e., the niche customers vanish and the mass customers expand), the service provider should inform them of the mean service quality only when the sample size is sufficiently large.

Proposition 4 contributes to the economics literature of quality disclosure (Dranove and Jin 2010, and references therein) by showing that customer bounded rationality alone can overturn the classic unraveling result that a *monopolistic* firm should disclose quality information whatever its quality is. This result is driven by the assumption of fully rational customers: Even a low-quality firm should disclose information because customers rationally infer non-disclosing firms as having the lowest quality. In our setting, however, customers cannot make this rational inference. Therefore, the role of information disclosure is not to signal quality but to influence the market composition (i.e., the niche & mass customers). A low-quality service provider should not disclose quality information because it reduces demand (i.e., the niche customers).

Managerially, Proposition 4 underscores the importance of an up-to-date understanding of customers' service quality information. Specifically, contrary to the conventional wisdom that a high- (low-) quality service provider always (never) discloses quality information, we find that when service quality is not too high or too low, lower-quality service providers may disclose information if

---

[4] As shown in the proof of Proposition 4, $\tilde{R}$ uniquely exists. Moreover, $\tilde{k} \in [1, +\infty)$ uniquely exists for $\underline{R} < R \leqslant \tilde{R}$ and does not exist otherwise.
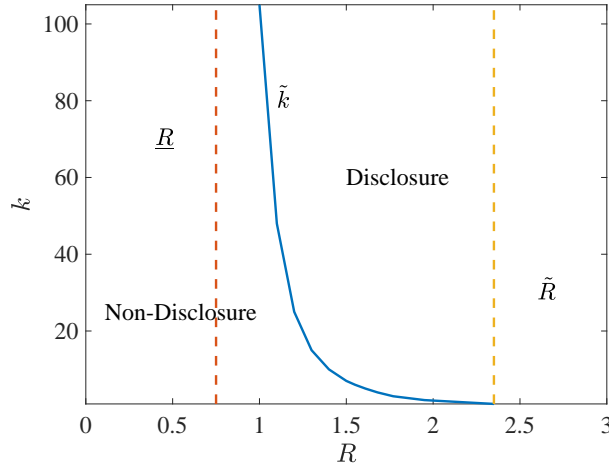
Figure 3   **The Impact of k and R on the Information Disclosure Decision ($\mu = 3, \lambda = 2, c = 1, \sigma = 5$, $\underline{R} = 0.75$,** $\tilde{R} = 2.346$**)**

word-of-mouth abounds in the market, whereas higher-quality service providers may not disclose if word-of-mouth is scarce. In addition, a manager who has chosen not to disclose information may still need to keep track of customers' word-of-mouth: As they share more information over time, he may switch to disclosing quality information.

Comparative statics. The next corollary characterizes the impact of service system parameters on the service provider's information disclosure decision; here we focus on the $\lambda < 2\mu$ case (since otherwise the server never discloses information).

COROLLARY 1. *When $\lambda < 2\mu$, $\tilde{k}$ strictly decreases in μ and strictly increases in c. When $\mu - \sqrt{c\mu/R} \leqslant \lambda < 2\mu$, $\tilde{k}$ strictly increases in λ.*

Corollary 1 again highlights the importance of incorporating customer bounded rationality in congested settings: Information non-disclosure (i.e., customers remain boundedly rational) is more attractive as the expected waiting cost increases (i.e., higher c, λ, or lower μ). In particular, compared to the no-congestion setting (i.e., $c = 0$), the server is incentivized to induce consumers to remain boundedly rational for a larger range of parameter values in congested settings. To see the intuition, first note that a higher expected waiting cost induces the service provider to reduce congestion by switching to targeting exclusively the niche customers. He does not disclose information because this reduces demand. Notably, we have numerically verified that $\tilde{k}$ strictly increases in λ even when $\lambda < \mu - \sqrt{c\mu/R}$. For completeness of the analysis, we have also examined the server's information disclosure decision under endogenous quality (see Online Supplement 2.3). We find that all results in the present section continue to hold qualitatively.

5. Other Modeling Considerations

In this section we extend the analysis in §4 by considering two modeling variations, i.e., welfare-maximizing service systems and customer heterogeneity in the sample size.

5.1. Welfare Maximization

Consider a social planner that takes control of a service system to maximize social welfare. Typical examples include regulators levying tolls on public services, such as access to public roads/bridges/tunnels, and passport/driver's license applications. As in §4, we characterize the social planner's pricing, quality control, and information disclosure decisions.

5.1.1. Pricing. The social planner's pricing decision is given by:

$$\max_{p \geqslant 0} \ W(p, k).$$

Consistent with Huang et al. (2013), we assume that customers are not compensated for receiving the service, i.e., $p \geqslant 0$. This prevents individuals with no need of the service from congesting the service system only for the compensation. Let $p_w^*(k)$ and $W_w^*(k)$ denote the optimal price and social welfare for a given sample size k, and let W_r^* denote the optimal social welfare in the fully rational benchmark. The next proposition provides a full characterization of $p_w^*(k)$ and $W_w^*(k)$.

PROPOSITION 5. *(i) When $R > \underline{R}$ and $k < \hat{k}_w$, the optimal price $p_w^*(k) = 0$ and $W_w^*(k) < W_r^*$, where $\hat{k}_w \equiv \left[\sigma \Phi^{-1} \left(\left(1 - \frac{\mu - \sqrt{c\mu/R}}{\lambda} \right)^+ \right) \right]^2 \Big/ \left(\sqrt{\frac{cR}{\mu}} - R \right)^2$.*

(ii) Otherwise, the optimal price $p_w^*(k) = R - \sqrt{\frac{cR}{\mu}} + \frac{\sigma \Phi^{-1} \left(1 - \frac{\mu - \sqrt{c\mu/R}}{\lambda} \right)}{\sqrt{k}}$ and $W_w^*(k) = W_r^*$. Moreover, the optimal price $p_w^*(k)$ strictly increases in k for $R > \frac{c\mu}{[(\mu - \lambda)^+]^2}$, strictly decreases in k for $R < \frac{c\mu}{[(\mu - \lambda)^+]^2}$, and is invariant in k for $R = \frac{c\mu}{[(\mu - \lambda)^+]^2}$.

Bounded rationality leads customers to deviate from the socially optimal joining rate, but does not influence joining customers' benefit from the service. Therefore, it never improves social welfare. To correct for the distorted joining rate, the social planner prices higher (lower) than the fully rational benchmark if the mean service quality is low (high). This induces the socially optimal joining rate (i.e., welfare loss does not exist) unless the mean service quality is high and customers are sufficiently boundedly rational. In this case, fully correcting for the distortion requires the social planner to set a negative price, which is infeasible. Therefore, he prices at zero and bounded rationality leads to welfare loss. Notice that in the no-congestion setting, the social planner always prices at zero for $R > 0$ since $\hat{k}_w \to +\infty$. Intuitively, in the absence of congestion, joining is always socially beneficial as long as the service quality is positive. This result highlights the vital role

of congestion in determining the impact of customer bounded rationality on managing service systems.

Compared to revenue maximization, Proposition 5 shows that bounded rationality affects welfare-maximizing service systems differently: It reduces the optimal revenue only when customers are sufficiently rational (i.e., $k > \tilde{k}$), whereas it reduces the optimal social welfare only when they are sufficiently boundedly rational (i.e., $k < \hat{k}_w$). The key insight is that the impact of bounded rationality on customers' joining behavior may have contrasting revenue and welfare implications. For example, when $R > \underline{R}$ and $k < \min\{\tilde{k}, \hat{k}_w\}$, bounded rationality leads customers to join much more than they should and thus harms social welfare. However, it increases the optimal revenue by allowing the service provider to raise price.

5.1.2. Quality control. Similar to §4.1.2, we incorporate the server's quality decision by assuming that he can choose the expected service quality R at cost aR^2. Therefore, his quality control problem is given by:

$$\max_{p, R \geqslant 0} W_R(p, R) = \lambda_a^k(p) R - \frac{c\lambda_a^k(p)}{\mu - \lambda_a^k(p)} - aR^2,$$

where the first two terms on the right-hand side represent social welfare and the last term is the quality investment. Let $\hat{p}_w$ and $\hat{R}_w$ denote the optimal price and quality. The following proposition characterizes the impact of k on $\hat{p}_w$ and $\hat{R}_w$ for sufficiently high market potential.[5]

PROPOSITION 6. *(i) A unique optimal price $\hat{p}_w$ exists.*

(ii) For sufficiently high market potential λ and $c > 0$, the optimal price $\hat{p}_w$ strictly decreases in k. The optimal quality $\hat{R}_w$ and social welfare are invariant in k when $a \leqslant \bar{a}$, and the social planner should not offer service when $a > \bar{a}$, where $\bar{a}$ uniquely exists.

The social planner uses the pricing decision to maximize social welfare for a given quality level, and uses the quality decision to balance between the maximized social welfare and the corresponding quality investment. When market potential is sufficiently high and quality investment is not too costly (i.e., $a \leqslant \bar{a}$), the service system is sufficiently congested and bounded rationality leads customers to join more than socially desirable. In response, the social planner prices higher than the fully rational benchmark and thus always achieves the first-best social welfare. As customers collect more anecdotes, they join less and the social planner prices lower to maintain the same joining rate and social welfare. This further implies that the trade-off between social welfare and

---

[5] We have analytically shown that the social planner should not offer service for sufficiently low market potential (see the proof of Proposition 6). Here we omit the presentation for brevity.

quality investment is unchanged. As a result, the social planner maintains the same service quality. Notably, this result is unique to congested settings. Without congestion, a larger sample size leads more customers to join as long as the service quality is positive. Therefore, the social planner should improve quality (see Online Supplement 2.2).

To complement Proposition 6, we have conducted a numerical study to examine the intermediate market potential case, as illustrated in Figure 4. In this situation, the service system is not too congested and bounded rationality leads customers to join less than socially desirable. When λ is moderately low, the joining rate is very low such that inducing the socially optimal joining rate requires a negative price. Since this is not feasible, the social planner prices at zero. As customers acquire more anecdotes, they join more and this higher service utilization drives the social planner to improve quality. However, when the service system is sufficiently congested, an even larger sample size leads the social planner to lower quality to alleviate congestion. When λ is moderately high, the joining rate is higher and the social planner can fully correct for customers' under-joining behavior at a positive price. As customers acquire more anecdotes, they join more and the social planner prices higher to maintain the first-best social welfare. However, he maintains the same quality level since the trade-off between social welfare and quality investment is the same.
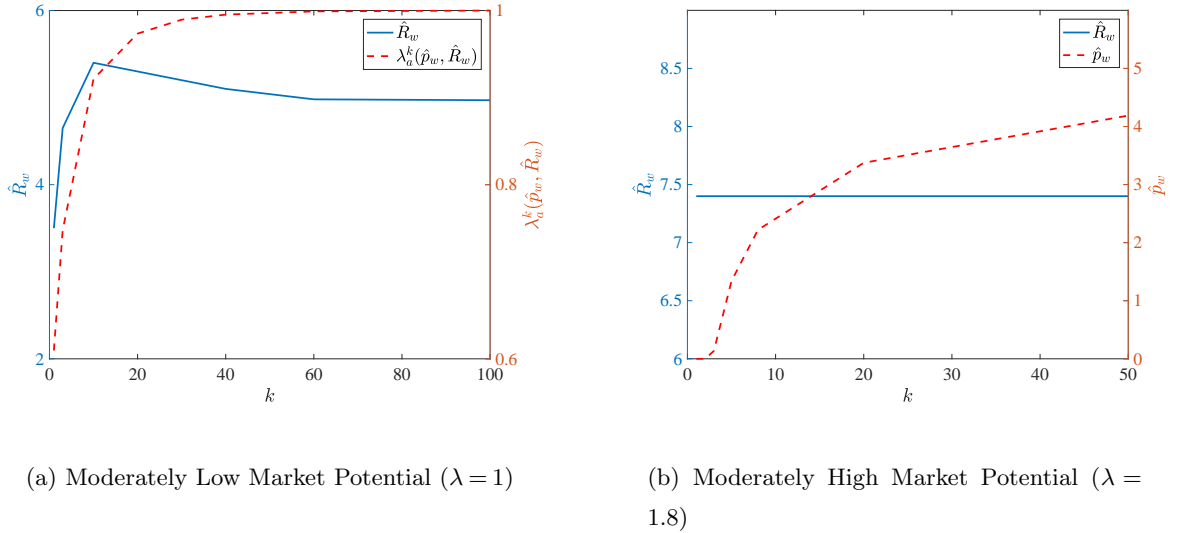

(a) Moderately Low Market Potential ($\lambda = 1$)

(b) Moderately High Market Potential ($\lambda = 1.8$)

Figure 4 **The Impact of k on $\hat{R}_w$, $\hat{p}_w$, and $\lambda_a^k(\hat{p}_w, \hat{R}_w)$ under Intermediate Market Potential ($a = 0.1, \mu = 2, c = 1, \sigma = 10$)**

5.1.3. Information disclosure. We now consider the social planner's decision to disclose service quality information. This has been widely practiced in the healthcare sector. For example, the Centers for Medicare & Medicaid Services (CMS 2017) in the US and the National Health Service (NHS 2016) in the UK disclose hospitals' key performance measures to the public online.

As in revenue maximization, we capture the information disclosure decision by assuming that the social planner chooses between informing customers of the mean service quality or not. Therefore, his joint pricing and information disclosure decisions are given by:

$$\max_{p \geqslant 0, i \in \{k, +\infty\}} W(p, i).$$

The next corollary, which follows immediately from Proposition 6, characterizes the information disclosure decision.

COROLLARY 2. *The service provider discloses information if $R > \underline{R}$ and $k < \hat{k}_w$, and is indifferent between disclosure and non-disclosure otherwise.*

Corollary 2 suggests that a high-quality public service manager should inform customers of the mean service quality when they are sufficiently boundedly rational. This is because bounded rationality leads to welfare loss in this case: It induces customers to join significantly less than socially desirable such that the manager cannot fully correct for the distortion with a non-negative price.

5.2. Heterogeneous Sample Sizes

To obtain sharper insights, we assumed in §4 that all customers estimate service quality based on k anecdotes. In practice, however, some customers may have a larger sample size than others because of easier access to anecdotes or higher cognitive capabilities. Consistent with the cognitive hierarchy model (Nagel 1995, Stahl and Wilson 1994), we incorporate this heterogeneity into our model by assuming that k follows a zero-truncated Poisson distribution with rate n.[6] Therefore, the proportion of customers with sample size i ($i = 1, 2, \ldots$) is equal to $f_i = \frac{n^i e^{-n}}{i!(1 - e^{-n})}$, where $n = \mathbf{E}[k]$ measures the average sample size across customers. Let λ_i denote the equilibrium joining rate of customers with i anecdotes, and let $\lambda_P^n \equiv \sum_{i=1}^{+\infty} f_i \lambda_i$ denote the total joining rate. We can show that

$$\lambda_i = \lambda \bar{\Phi} \left(\sqrt{i} \left(p + \frac{c}{\mu - \lambda_P^n} - R \right) / \sigma \right),$$

---

[6] We have truncated $k = 0$ from the Poisson distribution because the anecdotal reasoning framework requires that customers have at least one anecdote.

and then we have

$$\lambda_P^n = \lambda \sum_{i=1}^{+\infty} f_i \bar{\Phi}\left(\sqrt{i}\left(p + \frac{c}{\mu - \lambda_P^n} - R\right)/\sigma\right). \tag{5}$$

The next lemma shows that λ_P^n as defined by Equation (5) uniquely exists, where R_P is determined by $\sum_{i=1}^{+\infty} f_i \bar{\Phi}\left(\sqrt{i}\left[p + \frac{c}{(\mu - 0.5\lambda)^+} - R_P\right]/\sigma\right) = 0.5$.

LEMMA 3. *(i) A unique $\lambda_P^n \in (0, \min\{\lambda, \mu\})$ exists and strictly decreases in p.*

(ii) The joining rate λ_P^n strictly decreases in n for $R < R_P$, strictly increases in n for $R > R_P$, and is invariant in n for $R = R_P$.

(iii) $\lim_{n \to +\infty} \lambda_P^n(p) = \lambda_r(p)$.

Consistent with Lemma 1, Lemma 3 shows that a larger average sample size leads to fewer customers joining a low-quality service system and more customers joining a high-quality service system. In other words, incorporating customer heterogeneity in k does not qualitatively change the impact of customer bounded rationality on their joining behavior. Based on this, we have numerically verified that the impact of customer bounded rationality on the service provider's pricing, quality, and information disclosure decisions are also qualitatively preserved (see Online Supplement 3.2).

6. Concluding Remarks

In this paper, we studied the management of service systems with boundedly rational customers who infer service quality based on anecdotes. We characterized their equilibrium joining behavior and the service provider's pricing, quality, and information disclosure decisions. Bounded rationality induces customers to form heterogeneous service quality estimates; thus the service provider adopts a pricing strategy different from the fully rational benchmark. Specifically, a low-quality service provider targets the niche customers who considerably overestimate service quality, whereas a high-quality service provider targets the mass customers whose service quality estimates are not too low. With quality control, the service provider may reduce both price and quality as customers gather more anecdotes. We also characterized the service provider's quality information disclosure decision and found that it can be greatly influenced by customer bounded rationality: A lower-quality service provider may disclose information if customers are sufficiently rational, whereas a higher-quality service provider may not disclose if they are sufficiently boundedly rational.

This paper has several implications for service system management in practice. First, estimating customers' level of bounded rationality (i.e., the number of service quality anecdotes) is critical because it greatly influences a manager's optimal pricing, quality, and information disclosure decisions. We conjecture that ignoring customer bounded rationality (by assuming that customers

know the actual average service quality) can lead to significant profit loss. Second, even though conventional wisdom would probably suggest that managers should disclose quality information, we show that this may lower profit even for a high-quality service system. Third, when customers acquire more anecdotes over time, a service system manager may not improve service quality. In fact, when market potential is sufficiently high or low, he should reduce quality and compensate customers by lowering price. We hope that our theoretical model can stimulate future empirical studies in behavioral operations and help improve service management in practice.

For parsimony, we abstract away from several factors, which could serve as the basis for future modeling research. For example, we assumed that the service provider knows the service quality distribution. This is plausible for service systems with an established service content. However, it may fail if the service content is new, e.g., a restaurant experimenting with a new recipe. Future research can extend our work by considering service providers' decisions when they do not know the service quality distribution. Another research direction is to incorporate market competition in our model. It would be interesting to investigate whether customer anecdotal reasoning can soften service providers' competition on price and information disclosure. Finally, future research can consider customers' estimation of service quality based on both earlier customers' anecdotes and the service system itself (e.g., a prior belief, the queue length, or the expected waiting time).

References

Alizamir, S., F. de Véricourt, and P. Sun (2013). Diagnostic accuracy under congestion. *Management Sci. 59*(1), 157–171.

Anderson, E. W., C. Fornell, and R. T. Rust (1997). Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Sci. 16*(2), 129–145.

Ayers, B. (2011). How to get more Yelp reviews for your business. http://thegetsmartblog.com/how-to-get-more-yelp-reviews. Accessed on 23 November 2016.

CMS (2017). Find and compare information about hospitals: Hospital Compare. https://www.medicare.gov/hospitalcompare. Accessed on 23 February 2017.

Cui, S. and S. Veeraraghavan (2016). Blind queues: The impact of consumer beliefs on revenues and congestion. *Management Sci. 62*(12), 3656–3672.

Debo, L. and S. Veeraraghavan (2014). Equilibrium in queues under unknown service times and service value. *Oper. Res. 62*(1), 38–57.

Debo, L. G., C. Parlour, and U. Rajan (2012). Signaling quality via queues. *Management Sci. 58*(5), 876–891.

Dizik, A. (2013). Waiter, a pretty steak, please. http://www.wsj.com/articles/SB10001424127887324520904578553502283072828. Accessed on 20 March 2016.

Dranove, D. and G. Z. Jin (2010). Quality disclosure and certification: Theory and practice. *J. Econ. Lit. 48*(4), 935–963.

Edwards, W. (1968). *Conservatism in human information processing. In: Kleinmutz, B. (Ed.), Formal Representation of Human Judgement.* Wiley.

Fishman, M. J. and K. M. Hagerty (2003). Mandatory versus voluntary disclosure in markets with informed and uninformed customers. *J. Law, Economics & Organization 19*(1), 45–63.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *J. Law & Economics 24*(3), 461–483.

Guo, P., M. Haviv, and Y. Wang (2014). Equilibrium queueing strategies when service quality is unknown to some customers. Working Paper, The Hong Kong Polytechnic University, Hong Kong.

Guo, P., W. Sun, and Y. Wang (2011). Equilibrium and optimal strategies to join a queue with partial information on service times. *Eur. J. Oper. Res. 214*(2), 284–297.

Guo, P. and P. Zipkin (2007). Analysis and comparison of queues with different levels of delay information. *Management Sci. 53*(6), 962–970.

Guo, P. and P. Zipkin (2009). The effects of the availability of waiting-time information on a balking queue. *Eur. J. Oper. Res. 198*(1), 199–209.

Hassin, R. (1986). Consumer information in markets with random product quality: The case of queues and balking. *Econometrica. 54*(5), 1185–1195.

Hassin, R. (2007). Information and uncertainty in a queuing system. *Probab. Eng. Inform. Sci. 21*(03), 361–380.

Hassin, R. and M. Haviv (2003). *To queue or not to queue: Equilibrium behavior in queueing systems.* Springer.

Hirshleifer, D. A., S. S. Lim, and S. H. Teoh (2004). Disclosure to an audience with limited attention. *Available at SSRN 604142.*

Huang, T., G. Allon, and A. Bassamboo (2013). Bounded rationality in service systems. *Manufacturing Service Oper. Management 15*(2), 263–279.

Huang, T. and Y.-J. Chen (2015). Service systems with experience-based anecdotal reasoning customers. *Prod. Oper. Manag. 24*(5), 778–790.

Huang, T. and Y. Yu (2014). Sell probabilistic goods? A behavioral explanation for opaque selling. *Marketing Sci. 33*(5), 743–759.

Johnson, J. P. and D. P. Myatt (2006). On the simple economics of advertising, marketing, and product design. *Am. Econ. Rev. 96*(3), 756–784.

Jovanovic, B. (1982). Truthful disclosure of information. *Bell J. Econ. 13*(1), 36–44.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica. 47*(2), 263–291.

Kahneman, D. and A. Tversky (1984). Choices, values, and frames. *Am. Psychol. 39*(4), 341.

Kremer, M. and L. Debo (2015). Inferring quality from wait time. *Management Sci. 62*(10), 3023–3038.

Lahiri, A. and D. Dey (2013). Effects of piracy on quality of information goods. *Management Sci. 59*(1), 245–264.

Li, M., N. Petruzzi, and J. Zhang (2015). Overconfident competing newsvendors. *Management Sci. (Forthcoming)*.

Li, X., P. Guo, and Z. Lian (2016). Quality-speed competition in customer-intensive services with boundedly rational customers. *Prod. Oper. Manag. 25*(11), 1885–1901.

Mealey, L. (2016). How to price your restaurant menu. https://www.thebalance.com/how-to-price-your-restaurant-menu-2888593. Accessed on 23 January 2017.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *Bell J. Econ. 12*(2), 380–391.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *Am. Econ. Rev. 85*(5), 1313–1326.

NHS (2016). Find services – NHS Choices. https://www.nhs.uk/service-search. Accessed on 23 February 2017.

Nielsen (2013). Under the influence: Consumer trust in advertising. http://www.nielsen.com/us/en/insights/news/2013/under-the-influence-consumer-trust-in-advertising.html. Accessed on 20 March 2016.

Nielsen (2015). Digital formats are among the most trusted advertising sources despite slow growth. http://www.nielsen.com/id/en/insights/news/2015/digital-formats-are-among-the-most-trusted-advertising-sources-despite-slow-growth.html. Accessed on 20 March 2016.

Osborne, M. J. and A. Rubinstein (1998). Games with procedurally rational players. *Am. Econ. Rev. 88*(4), 834–847.

Power, D. (2011). How to use Yelp Events to win new customers. http://sproutsocial.com/insights/how-to-use-yelp-events-new-customers. Accessed on 23 November 2016.

SAMPi (2015). Wechat marketing Part II: Getting followers organically. http://sampi.co/wechat-marketing-part-ii-getting-followers-organically. Accessed on 20 March 2016.

Spiegler, R. (2006a). Competition over agents with boundedly rational expectations. *Theor. Econ. 1*(2), 207–231.

Spiegler, R. (2006b). The market for quacks. *Rev. Econ. Stud. 73*(4), 1113–1131.

Stahl, D. O. and P. W. Wilson (1994). Experimental evidence on players' models of other players. *J. Econ. Behav. Organ. 25*(3), 309–327.

Stivers, A. E. (2004). Unraveling of information: Competition and uncertainty. *Top. Theor. Econ. 4*(1), 1–18.

Sun, M. (2012). How does the variance of product ratings matter? *Management Sci. 58*(4), 696–707.

Szech, N. (2011). Becoming a bad doctor. *J. Econ. Behav. Organ. 80*(1), 244–257.

Wang, X., L. G. Debo, A. Scheller-Wolf, and S. F. Smith (2010). Design and analysis of diagnostic service centers. *Management Sci. 56*(11), 1873–1890.

Werner, L. (2014). A new app gets diners rewards for posting their photos on Facebook. http://www.forbes.com/sites/lauriewerner/2014/07/29/a-new-app-gets-diners-rewards-for-posting-their-photos-on-facebook/#72ee6b7a4afd. Accessed on 20 March 2016.

Wolff, R. W. (1982). Poisson arrivals see time averages. *Oper. Res. 30*(2), 223–231.