Prioritization and Price-Plus-Delay Competition with Self-Selecting, Heterogeneous, Time-Sensitive Customers

Arvind Sainathan

Prioritization and Price-Plus-Delay Competition with Self-Selecting, Heterogeneous, Time-Sensitive Customers

Time is often used as a differentiating factor in several service operations contexts by service providers (SPs) who prioritize their customers. We investigate the performance of such differentiated service vis-a-vis *single service* in which customers are not prioritized, and analyze how it is affected by customers' self-selection and competition. Customers are of two types with different delay sensitivities: *impatient* and *patient*. An SP providing differentiated service offers two service classes. We first establish *strong* incentive compatibility (IC) conditions that this SP needs to satisfy; otherwise, the queuing dynamics results in all the customers selecting just a single service class. We show that these conditions are stronger than IC conditions which only *preclude customers from changing their choices after they have been made*. We then derive multiple results pertaining to different equilibriums under monopoly and duopoly. Two of them are especially noteworthy: (i) under monopoly, we identify a novel criterion for differentiated service to do better: *sufficient customer heterogeneity*; (ii) under duopoly, *customer composition* also becomes important due to competition between the two SPs: specifically, we find that if the fraction of impatient customers is high, *the SPs do not prioritize customers at equilibrium and they are better from doing that.*

Keywords: service operations, duopoly, incentive compatibility, heterogeneity, prioritization

1. Introduction

Differentiated service offerings, in which some customers are prioritized, are often used by service providers (SPs) to differentiate customers who are heterogeneous in their delay sensitivities. Some examples include (i) amusement/theme parks which allow customers to *jump queues* by purchasing high-priced priority tickets and have faced some backlash from the practice (Wallop 2010), (ii) *concierge medicine* in which physicians/general practitioners provide faster service and patients pay a premium in comparison to regular primary care (e.g., see Wieczner (2013) and Gavirneni and Kulkarni (2014)), and (iii) restaurants charging customers additional fees upfront when they make reservations which enable them to avoid waiting for tables later (e.g., see Knowledge@Wharton (2013) and Markovich (2014) who discuss about reservation fees). Although these examples might suggest that offering differentiated service is always better than offering *single service* in which there is no prioritization and all customers are seen on a first-come-first-serve (FCFS) basis, that is not true in practice. For instance, there are some theme parks that do not have priority tickets, physicians who do not offer concierge medicine, and restaurants that do have any reservations. We analyze when a SP benefits from offering differentiated service instead of single service by focusing on three key aspects: *self-selecting customers*, *heterogeneity in their delay sensitivities*, and *competition between SPs*.

We consider two types of customers, impatient and patient, that differ in their delay sensitivities. They self-select and make their choice based on the price(s) charged by SP(s) and the delay that they would expect at equilibrium. An SP either (i) offers single service and charges a single price for it, or (ii) offers differentiated service with two service classes in which she charges different prices for *high-priority* and *low-priority* services. The SP also anticipates how customers would make their choices at equilibrium, and optimizes her pricing and service delivery decisions. We first show that the prices under differentiated service have to satisfy *strong* incentive compatibility (IC) conditions; otherwise, all customers are better from selecting either high-priority or low-priority service. We find that these conditions are stronger than IC conditions which prevent customers from changing their choices after they have been made already. In order to better understand the impact of competition, we analyze the optimal service delivery of SP(s) under two cases: (i) *monopoly* with a single SP and (ii) *duopoly* in which two SPs engage in *price-plus-delay competition*.

In a monopoly, when customers highly value the service, we find that differentiated service is beneficial only if the customers are *sufficiently heterogeneous*. This result marks a significant departure from the research literature in queuing optimization and service operations management,

which mostly assumes (implicitly or explicitly) that offering differentiated service through prioritization would be better than offering single service[1]. On the contrary, we find that the presence of self-selecting customers and the resulting strong IC conditions can make differentiated service worse than single service. Hence, we identify a key feature that needs to be considered by the SP in making the service delivery decision: *heterogeneity in customers' delay sensitivities*[2].

In a duopoly, we consider equilibriums involving the SPs' pricing decisions, resulting from the price-plus-delay competition between them, under three different types of service deliveries: (i) both SPs provide single service, (ii) one of them provides single service and the other provides differentiated service, and (iii) both of them provide differentiated service. In the first case, we find that there can be two kinds of equilibrium based on the composition of customers: *low-price equilibrium* when patient customers comprise a majority and *high-price equilibrium* when impatient customers form a majority. We also characterize the unique symmetric equilibrium in the third case. We compare the three cases and find the *equilibrium service delivery* of the SPs for different numerical examples. In particular, we find that *if the fraction of impatient customers is high then the SPs are better from offering single service and the resulting high-price equilibrium.* Hence, in a duopoly, *customer composition* is also a key feature that has to be taken into account by an SP while deciding how to deliver the service.

2. Literature Review

The research in this paper is related to the literature on queuing optimization and equilibrium consumer behavior in queuing systems. Hassin and Haviv (2003) examine this literature in detail. Among the research that they review, this paper is closely related to Lederer and Li (1997). The model here is similar to that in Lederer and Li (1997) who also consider time-sensitive customers and price-plus-delay competition between firm. However, they assume that the number of firms is so large that the "full price" (price plus waiting cost) at equilibrium does not depend on the prices charged by the firms. We do not make this assumption, which fundamentally alters the nature of the problem and leads to different results. An important implication is as follows. Lederer and Li (1997) mainly consider (preemptive) priority queues because they are optimal for the firms in their setting. However, we find that there are scenarios in which prioritization is sub-optimal because (i) price charged by a firm affects the full price in our model so the firm decides it strategically,

[1] For details, see §2.

[2] A lack of capacity may also result in differentiated service being infeasible and hence worse than single service (Sainathan 2014). However, we show that even if capacity is sufficient (and customers highly value the service), differentiated service is still worse if they are not sufficiently heterogeneous.

and (ii) customers self-select. Armony and Haviv (2003) also comes close to this paper because they consider a duopoly facing demand from two customer classes with different delay sensitivities. However, they assume that all customers are served on a first-come-first-served basis, and do not consider priority queues. In this regard, whether a service provider should prioritize or not is a key question we analyze in the paper, and which neither Lederer and Li (1997) nor Armony and Haviv (2003) consider.

Other research works consider consumer behavior in conjunction with pricing in the context of service operations management and are related to this paper. Luski (1976) was one of the pioneers to consider price-plus-delay competition in a duopoly and finds scenarios when identical firms charge different prices at equilibrium. However, he does not consider prioritization or heterogeneity in delay sensitivity among customers. Li and Lee (1994) also consider a similar setting but with non-identical firms and possible jockeying among customers, and show that the faster firm always charges a higher price at equilibrium. Li et al. (2012) extend the analysis to naive customers who do not know the firms' service rates but observe their queue lengths and use that information to make the decisions. Chen and Wan (2003) study the competition between two make-to-order firms in which the firms may provide service with different value and customers may have firm-dependent waiting costs. Some research papers have looked at pricing along with priority queues and/or customer heterogeneity. Rao and Peterson (1998) analyze the pricing of priority services for a service center with a fixed finite number of customers who maximize their own profits. Zhang et al. (2007) consider the optimal pricing of communication services for customers with different service valuations but same delay sensitivity. Afeche (2013) and Katta and Sethuraman (2005) model heterogeneity in both service valuation and delay sensitivity among customers. They optimize the service provider's revenue subject to individual rationality (IR) and incentive compatibility (IC) conditions among different customer classes. However, they consider only prioritization because they find it to be always optimal, and do not analyze competition between service providers. Hsu et al. (1998) examine optimal pricing and scheduling subject to IC conditions in the presence of different quality-of-service guarantees for multiple customer classes, and show that randomized prioritization may be optimal. Afeche et al. (2013) model risk aversion in customers and analyze how time-sensitive services should be priced then. Pangburn and Stavrulaki (2008) investigate joint pricing and capacity decisions for a firm facing heterogeneous customers who are dispersed among different locations. It has to decide between offering segmented and pooled services. Segmented service examined by them is different from prioritized service in this paper in one key aspect: there is no loss in economies of scale from prioritization because the entire capacity is used. Furthermore,

they do not model competition. As in this paper, Sainathan (2014) also compare differentiated service with *single service*. However, he considers an SP offering ancillary service in which all the customers have to satisfied while minimizing the cost of service provision.

A central feature of this paper is the presence of *strong* IC conditions (see §3). Some research papers have analyzed optimal pricing in the context of service operations while ignoring IC constraints either because they consider aggregate demand (instead of individual customer choices) or they assume that the firm allocates customers who do not self-select. Some examples include So and Song (1998), Boyaci and Ray (2003), Allon and Federgruen (2009), and Anand et al. (2011).

To the best of our knowledge, this paper is the first research to view prioritization as a service provider's *strategic choice*. Most prior research either take prioritization for granted (because it obviously performs better) or do not consider it at all[3]. However, this paper compares differentiated service (through prioritization) with single service, and analyzes how two key aspects, *heterogeneity in customers' delay sensitivities* and *competition between service providers*, affect the SP's decision to prioritize.

3. Model

The arrival of customers for service is a Poisson process with rate λ. Customers are of two types, *impatient* (Type 1) and *patient* (Type 2), with *delay sensitivities* given by η_1 and η_2 ($\eta_1 > \eta_2 > 0$) respectively. Due to waiting, they experience delay costs given by the products of delay sensitivity and *waiting time*[4]. A customer selects a service provider (SP) based on the *dis-utility*, which is the sum of the price charged by SP and the expected delay cost incurred by him. If the dis-utility does not exceed the valuation v, the customer buys the service; otherwise, he does not buy it. The fraction of impatient (patient) customers is q $(1-q)$. Without loss of generality, SP's capacity (the rate at which she provides service) is normalized to one and $0 < \lambda < 1$. An SP provides service in one of the following two ways: (i) *single service* in which the customers are served on a first-come-first-serve (FCFS) basis and all of them pay the same price, and (ii) *differentiated service* (with two service classes) in which some customers are prioritized [5] and the *high-priority service* is charged a higher price than the *low-priority service*. If the SP provides single service, the expected waiting

[3]Sainathan (2014) do compare prioritized and single services while Pangburn and Stavrulaki (2008) compare segmented and pooled services. However, as we explain above, their settings are different. Importantly, Sainathan (2014) analyze ancillary services in which cost is minimized and single service becomes optimal due to low capacity, while segmented service analyzed in Pangburn and Stavrulaki (2008) is different from prioritized service in this paper.

[4]We define waiting time as the sum of time in the queue and service time.

[5]We do not consider strategic delay because we focus on scenarios involving significant face-to-face interaction between customers and SPs in which implementing it would be difficult. That is because if customers get to know that they are delayed purposely, then SPs might face a significant backlash from them.

time is $W(\delta)$ in which δ is the fraction of total customers that purchases the service from her. We assume that $v > \eta_1 W(0)$, i.e., if the price is zero and there is no queuing at the SP, any customer would buy the service. If she provides differentiated service then we let δ_h (δ_l) be the fraction of total customers that purchase the high-priority (low-priority) service from her. The expected waiting times for high-priority and low-priority services are then given by $W_h(\delta_h)$ and $W_l(\delta_h, \delta_l)$ respectively. We define a customer's *net utility* as the difference of his valuation and dis-utility. A central feature of our model is the self-selecting behavior of customers. Next, we discuss about how they do this self-selection and the implications it has on the SP's pricing decisions.

We first consider a single SP. The customers either purchase from her or do not buy anything. If she provides a single service at price p there are four possibilities at equilibrium: (i) $p+\eta_1 W(1) \leq v$, then $\delta = 1$ and all the customers purchase the service; (ii) $p + \eta_1 W(1-q) < v < p + \eta_1 W(1)$ then $1 - q < \delta = W^{-1}\left(\frac{v-p}{\eta_1}\right) < 1$, some impatient customers and all patient ones purchase the service; (iii) $p + \eta_2 W(1-q) \leq v \leq p + \eta_1 W(1-q)$ then $\delta = 1 - q$, none of the impatient customers buys from the SP but all patient ones do; and (iv) $v < p + \eta_2 W(1-q)$ then $\delta = W^{-1}\left(\frac{v-p}{\eta_2}\right) < 1 - q$ and only some patient customers purchase from the SP. Note that although customers are served on an FCFS basis, patient customers, due to their lower delay sensitivity, *get a preference* over impatient ones in terms of who gets served. If the SP provides differentiated service, she charges prices p_h and p_l for high-priority and low-priority services respectively. Also, $\delta_h, \delta_l > 0$; otherwise, she can provide just a single service. The prices should satisfy incentive compatibility (IC) conditions for δ_h and δ_l to be the fractions purchasing high-priority and low-priority services at equilibrium. These conditions are given by

$$p_h - p_l \leq \eta_1 \left(W(\delta_h + \delta_l) - W_h(0)\right), \tag{1}$$

$$p_h - p_l \geq \eta_2 \left(W_l(\delta_h + \delta_l, 0) - W(\delta_h + \delta_l)\right). \tag{2}$$

We provide the reasoning for these conditions as follows. Suppose (1) does not hold, then $p_h - p_l > \eta_1 \left(W(\delta_h + \delta_l) - W_h(0)\right) > \eta_2 \left(W(\delta_h + \delta_l) - W_h(0)\right)$ since $W(\delta_h + \delta_l) > W_h(0)\ \forall \delta_h, \delta_l > 0$. Hence, the selection of low-priority service by all the $\delta_h + \delta_l$ fraction of customers becomes an equilibrium. Further, it's a *Pareto dominating* equilibrium because every customer gets a higher net utility from selecting low-priority service than that obtained from choosing high-priority service. Similarly, if (2) does not hold then $p_h - p_l < \eta_2 \left(W_l(\delta_h + \delta_l, \delta_h + \delta_l) - W(\delta_h + \delta_l)\right) < \eta_1 \left(W_l(\delta_h + \delta_l, \delta_h + \delta_l) - W(\delta_h + \delta_l)\right)$. Then the selection of high-priority service by all the $\delta_h + \delta_l$ fraction of customers becomes a Pareto dominating equilibrium. Now consider two other conditions

that are given as follows:

$$p_h - p_l \leq \eta_1 \left(W_l(\delta_h, \delta_l) - W_h(\delta_h)\right), \tag{3}$$

$$p_h - p_l \geq \eta_2 \left(W_l(\delta_h, \delta_l) - W_h(\delta_h)\right). \tag{4}$$

Condition (3) (Condition (4)) ensures that impatient (patient) customers are not worse from selecting high-priority (low-priority) service instead of low-priority (high-priority) service. Therefore, for prices satisfying (3) and (4), selection of high-priority and low-priority services by customer fractions δ_h and δ_l respectively is an equilibrium, provided (i) $\delta_h \leq q$ and $\delta_l \leq 1 - q$, and (ii) the dis-utilities from high-priority and low-priority services do not exceed the valuation v. *However, if condition (1) or (2) is violated, it is not a Pareto dominating equilibrium.* In this case, there is another equilibrium in which customers are better by all of them selecting either high-priority (if (2) is violated) or low-priority service (if (1) is violated). Next, we characterize in Proposition 1 the relationship between IC conditions (1) and (2) and conditions (3) and (4). All proofs are in the Appendix.

Proposition 1 *Under both $M/M/1$ preemptive and non-preemptive priority, $W_l(x, y-x) - W_h(x)$ is strictly increasing in x $\forall x \in [0, y], y > 0$. Hence, (1) $\Rightarrow$ (3) and (2) $\Rightarrow$ (4).*

Proposition 1 implies that if having more customers selecting the high-priority service (while the total amount of customers remains the same) has a higher impact on low-priority service, i.e., it increases W_l more than it increases W_h (which is true for $M/M/1$ preemptive and non-preemptive priority queuing systems), then the IC conditions given by (1) and (2) *are stronger* than the conditions in (3) and (4) respectively. Because customers are self-selecting, we assume that they are utility-maximizing and so a Pareto dominating equilibrium is more likely. Hence, if the SP provides prioritized service, we require that the prices satisfy the *strong* IC conditions given by (1) and (2) so that there are some customers selecting each of high-priority and low-priority services. Next, we find how these two conditions affect customers' service choices.

Proposition 2 *Under both $M/M/1$ preemptive and non-preemptive priority, the IC conditions (1) and (2) imply that impatient customers select only high-priority service and patient ones select only low-priority service.*

Proposition 2 shows that the prices satisfying IC conditions (1) and (2) result in a *perfect differentiation* of customers in which impatient and patient customers select high-priority and low-priority services respectively. It is important to note that this feature is not an assumption of our model

but it is a result of the customers' self-selecting behavior. The SP sets the prices *sufficiently different*, so that (1) and (2) are satisfied, in order to preclude the entire purchase resulting from just the high-priority or low-priority service under a Pareto-dominating equilibrium. That results in a perfect differentiation of customers by the SP. Next, we consider what happens with two SPs.

There are three possible ways in which service can be provided with two SPs: (i) each of them provides single service, (ii) one of them provides single service while the other prioritizes customers and provides differentiated service, and (iii) each of them prioritizes customers and provides differentiated service. We consider each of these cases in detail in §5. With two SPs, we further note that (i) the price(s) charged by an SP are influenced by the price(s) charged by the other SP due to *price-plus-delay* competition between them, and (ii) if an SP prioritizes her customers the prices should still satisfy the strong IC conditions given by (1) and (2).

For analytical conciseness, in the rest of the paper, we model SP(s) with an $M/M/1$ queuing system and having a preemptive priority for high-priority service under differentiated service. So we have $W(\delta) = W_h(\delta) = 1/(1-\lambda\delta)$ and $W_l(\delta_h, \delta_l) = 1/\left((1-\lambda\delta_h)\cdot(1-\lambda\delta_h-\lambda\delta_l)\right)$. We first analyze when prioritizing customers is optimal under a monopolistic SP.

4. Monopoly

The SP can provide two types of services: single service and differentiated service. We first consider single service, formulate the SP's problem, and characterize her optimal solution.

4.1 Single Service

The SP's profit maximization problem can be formulated as

$$
\begin{aligned}
(\mathcal{S}) \qquad & \max_{\delta,p} \pi_S = \lambda\delta p \\
s.t.\ (\delta+q-1)^+\left(p+\frac{\eta_1}{1-\lambda\delta}\right) &\leq (\delta+q-1)^+ v \qquad (5)\\
p+\frac{\eta_2}{1-\lambda\delta} &\leq v \qquad (6)\\
0\leq\delta\leq 1; p &\geq 0.
\end{aligned}
$$

The SP can provide service in two ways: (i) she sets the price high so that only patient customers purchase the service, $\delta \leq 1-q$ here and so constraint (5) is automatically satisfied, or (ii) she sets the price low to satisfy all the patient customers and some impatient customers so that $\delta > 1-q$ and constraint (5) implies that the net utility of these impatient customers from purchasing the service is non-negative. Constraint (6) is always satisfied because the retailer is profitable ($v > \eta_2 W(0) = \eta_2$

ensures that) and patient customers *get a preference* over impatient ones due to their lower delay sensitivity and customers' self-selection. Next, we characterize the optimal solution of $\mathcal{S}$.

Proposition 3 *Under optimality, either (5) or (6) binds. The optimal values are given as follows:*

1. *If* $\sqrt{\frac{\eta_2}{v}} > 1-\lambda(1-q)$ *then* $\delta^* = \frac{1}{\lambda}\left(1-\sqrt{\frac{\eta_2}{v}}\right)$, $p^* = v - \frac{\eta_2}{1-\lambda\delta^*}$, *and* $\pi_S^* = \left(\sqrt{v}-\sqrt{\eta_2}\right)^2$.
2. *If* $\sqrt{\frac{\eta_2}{v}} \leq 1-\lambda(1-q) \leq \sqrt{\frac{\eta_1}{v}}$ *then* $\delta^* = 1-q$, $p^* = v - \frac{\eta_2}{1-\lambda(1-q)}$, *and* $\pi_S^* = \lambda(1-q)\left(v - \frac{\eta_2}{1-\lambda(1-q)}\right)$.
3. *If* $\sqrt{\frac{\eta_1}{v}} < 1-\lambda(1-q)$ *then let* $\hat{\delta} \equiv \min\left(\frac{1}{\lambda}\left(1-\sqrt{\frac{\eta_1}{v}}\right), 1\right)$. *If* $\lambda\hat{\delta}\left(v - \frac{\eta_1}{1-\lambda\hat{\delta}}\right) \geq (<)$ $\lambda(1-q)\left(v - \frac{\eta_2}{1-\lambda(1-q)}\right)$ *then* $\delta^* = \hat{\delta}$ $(\delta^* = 1-q)$, $p^* = v - \frac{\eta_1}{1-\lambda\delta^*}$ $\left(p^* = v - \frac{\eta_2}{1-\lambda(1-q)}\right)$, *and* $\pi_S^* = \eta_1\frac{(\lambda\delta^*)^2}{(1-\lambda\delta^*)^2}$ $\left(\pi_S^* = \lambda(1-q)\left(v - \frac{\eta_2}{1-\lambda(1-q)}\right)\right)$.

Proposition 3 shows that the optimal price and fraction of customers who buy the service depend on the ratios of customers' delay sensitivities and their valuation. If these ratios are high (for both patient and impatient customers) then the SP serves *some patient customers*. However, if the ratio is low for patient but high for impatient customers, then the SP serves *all the patient customers*. In both these cases, serving any impatient customer is unprofitable for the retailer. Finally, if the ratios are low (for both patient and impatient customers), the SP is faced with two choices. She can either (i) charge a high price and serve all patient customers but none of the impatient ones or (ii) charge a low price and serve all the patient customers and some/all of the impatient ones. The better choice is determined based on whether the increase in demand (from serving impatient customers) is able to compensate for the price reduction. Next, we consider the SP providing differentiated service.

4.2 Differentiated Service

The SP's profit maximization problem can be formulated as

$$
\begin{aligned}
(\mathcal{D}) \qquad & \max_{\delta_h,\delta_l,p_h,p_l} \pi_D = \lambda\left(\delta_h p_h + \delta_l p_l\right) \\
s.t. \qquad p_h + \frac{\eta_1}{1-\lambda\delta_h} &\leq v && (7)\\
p_l + \frac{\eta_2}{(1-\lambda\delta_h)\cdot(1-\lambda\delta_h-\lambda\delta_l)} &\leq v && (8)\\
p_h - p_l &\leq \eta_1\left(\frac{\lambda\left(\delta_h+\delta_l\right)}{1-\lambda\delta_h-\lambda\delta_l}\right) && (9)\\
p_h - p_l &\geq \eta_2\left(\frac{\lambda\left(\delta_h+\delta_l\right)}{\left(1-\lambda\delta_h-\lambda\delta_l\right)^2}\right) && (10)\\
0 \leq \delta_h \leq q; 0 \leq \delta_l &\leq 1-q; p_h, p_l \geq 0. && (11)
\end{aligned}
$$

Constraints (7) and (8) ensure that impatient and patient customers, who purchase high-priority and low-priority services respectively (see Proposition 2), obtain non-negative net utilities. Constraints (9) and (10) correspond to IC conditions (1) and (2). They make high-priority and low-priority services incentive compatible in the presence of self-selecting customers. Finally, constraints in (11) require that (i) fraction of customers that select high-priority (low-priority) service does not exceed the fraction of impatient (patient) customers and (ii) prices be non-negative. We let $\delta_t \equiv \delta_h + \delta_l$ to denote the total fraction of customers who buy high-priority and low-priority services. Next, we characterize the optimal solution of $\mathcal{D}$.

Proposition 4 *Under optimality, constraints (7) and (10) are binding, and there exists a unique* δ_t^* *that maximizes the SP's profit under differentiated service. Further, the other optimal values are given by* $\delta_h^* = \min\left(\max\left(0, \delta_t^* + q - 1, \frac{\sqrt{\eta_2} - (1-\lambda\delta_t^*)\sqrt{\eta_1}}{\lambda\sqrt{\eta_2}}\right), q\right)$, $\delta_l^* = \delta_t^* - \delta_h^*$, $p_h^* = v - \frac{\eta_1}{1-\lambda\delta_h^*}$, *and* $p_l^* = v - \frac{\eta_1}{1-\lambda\delta_h^*} - \frac{\eta_2\lambda\delta_t^*}{(1-\lambda\delta_t^*)^2}$, *and* δ_t^* *and* δ_h^* *are increasing in the valuation* v.

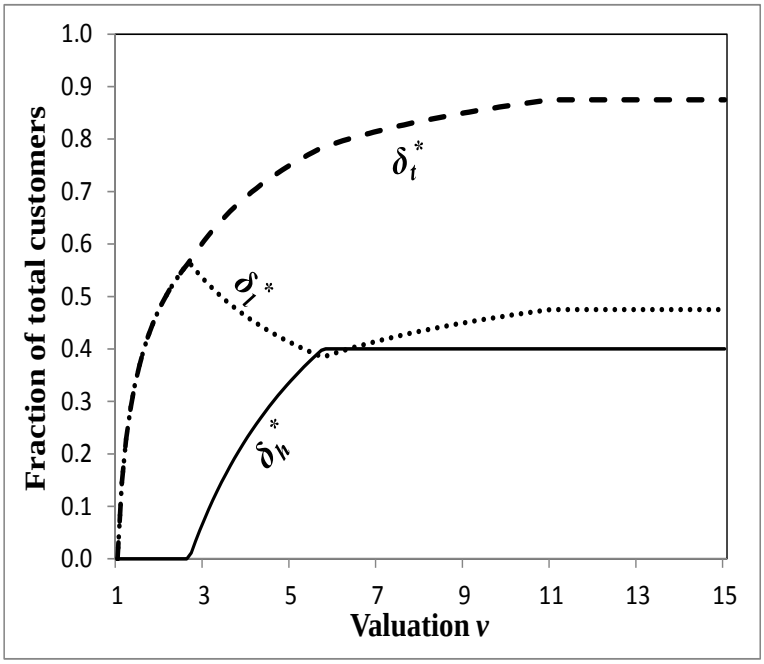

Figure 1: Variation of δ_h^*, δ_l^*, and δ_t^* with v when $\lambda = 0.8$, $\eta_1 = 1$, $\eta_2 = 0.3$, and $q = 0.4$

The upper bound on δ_t^* in Proposition 4 ensures that the IC constraints (9) and (10) can be satisfied under optimality. If the customers are not *sufficiently heterogeneous* so that $(1 - \eta_2/\eta_1)/\lambda < 1$ then *the self-selecting behavior of customers limits how much of them can be served by the SP providing differentiated service.* If too many customers purchase from the SP then, depending on prices p_h and p_l, all of them either select the high-priority or low-priority service. The SP does not

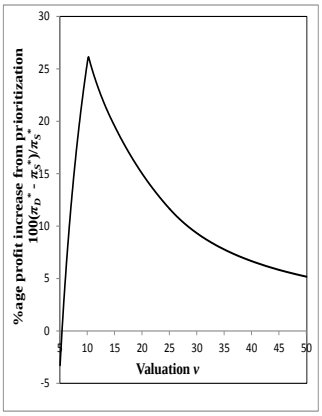

Figure 2: Low η_2: Variation of percentage benefit from prioritization when $\lambda = 0.8$, $\eta_1 = 1$, $\boldsymbol{\eta_2 = 0.1}$, and $q = 0.4$

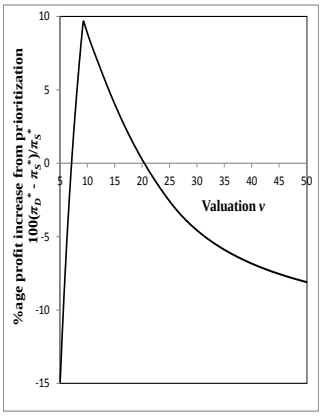

Figure 3: Intermediate η_2: Variation of percentage benefit from prioritization when $\lambda = 0.8$, $\eta_1 = 1$, $\boldsymbol{\eta_2 = 0.3}$, and $q = 0.4$

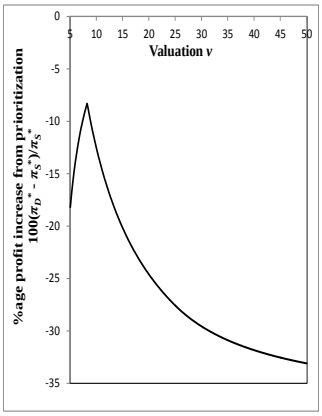

Figure 4: High η_2: Variation of percentage benefit from prioritization when $\lambda = 0.8$, $\eta_1 = 1$, $\boldsymbol{\eta_2 = 0.5}$, and $q = 0.4$

prefer that because it defeats the purpose of providing differentiated service. Next, we consider how the optimal fractions δ_h^*, δ_l^*, and δ_t^* change with customers' service valuation v.

Figure 1 illustrates the results from Proposition 4 in that the fractions δ_h^* and δ_t^* are monotonically increasing in v. Initially, when v is low, $\delta_h^* = 0$ and none of the impatient customers purchase from the SP. The SP would rather sell just to the patient customers (and charge them more) instead of offering some impatient customers high-priority service (and charging less for the impatient customers). When v takes intermediate values, δ_h^* increases from zero to q, the fraction of impatient customers. When v is high, all the impatient customers are satisfied. The total fraction of customers purchasing from the SP increases from zero to the upper bound $(1 - \eta_2/\eta_1)/\lambda = 0.875$. Finally, Figure 1 shows that the variation of δ_l^* with respect to v is *non-monotone*. Initially, when v is low, $\delta_l^* = \delta_t^*$ and increases in v. However, later when v takes intermediate values, we find that δ_l^* is *decreasing* in v (even though δ_h^* and δ_t^* are increasing). This result is explained as follows: *the SP benefits from more proportion of its customers selecting the high-priority service which is priced at a premium in comparison to the low-priority service.* When v is high, the high-priority service satisfies all the impatient customers and so δ_l^* increases in v until it's limited by the upper bound on the total fraction of customers served by the SP. Next, we compare the performances of single service and differentiated service.

4.3 Single Service vs. Differentiated Service

We compare π_S^* and π_D^*, the optimal profits from single service and differentiated service respectively. Next, we characterize how they change with v and how they are related to each other.

Theorem 1 *Both* π_S^* *and* π_D^* *are convex and strictly increasing functions of* v*. There are three possibilities: (i)* $\exists\ \underline{v}$ *s.t.* $\pi_D^* \leq \pi_S^*\ \forall v \leq \underline{v}$ *and* $\pi_D^* > \pi_S^*$ *otherwise; (ii)* $\exists\ \underline{v}$ & $\overline{v}$ *with* $\underline{v} < \overline{v} < \infty$ *s.t.* $\pi_D^* \leq \pi_S^*\ \forall v \leq \underline{v}$ & $\forall v \geq \overline{v}$ *and* $\pi_D^* > \pi_S^*$ *otherwise; and (iii)* $\pi_D^* \leq \pi_S^*\ \forall v$*. Further, if* $\frac{1}{\lambda}\left(1 - \frac{\eta_2}{\eta_1}\right) < (\geq)\ 1$ *then* $\lim_{v\to\infty} \pi_D^* < (>)\ \pi_S^*$*.*

Figures 2-4 illustrate the three possibilities in Theorem 1. In all the three figures, when the valuation v is low, differentiated service performs worse than single service. That is because the low valuation of service by customers makes the SP's capacity insufficient for prioritizing them well. However, for higher values of v, the value of η_2 also determines how differentiated service performs vis-a-vis single service. For these values of v, differentiated service performs better than single service in Figure 2 in which η_2 is low while the opposite is true in Figure 4 with a high value of η_2. In Figure 3, in which η_2 has an intermediate value, we find that differentiated service performs better when v takes intermediate values while single service performs better when v is high. Finally, we observe that the sub-optimality of differentiated service under very high values of v (in Figures 3 and 4) is driven by the *strong* IC constraints (9) and (10) which, in turn, result from the self-selecting behavior of customers. We provide the reasoning as follows. Suppose the SP only has to satisfy the *weak* IC conditions in (3) and (4) which, under $M/M/1$ priority, become $p_h - p_l \leq \eta_1\left(\frac{\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)}\right)$ and $p_h - p_l \geq \eta_2\left(\frac{\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)}\right)$ respectively. Then it can be shown under optimality that $p_h^* = v - \frac{\eta_1}{1-\lambda\delta_h^*}$ and $p_l^* = v - \frac{\eta_1}{1-\lambda\delta_h^*} - \frac{\eta_2\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)}$[6]. Further, for very high values of v, all the customers would be satisfied under both single and differentiated services. Therefore, after some algebra, we find that the profit under differentiated service (if only the weak IC conditions are satisfied) would have been $v - \frac{\eta_1}{1-\lambda} + \frac{(\eta_1-\eta_2)\lambda(1-q)}{(1-\lambda q)(1-\lambda)} > v - \frac{\eta_1}{1-\lambda}$, which is the profit under single service. So the presence of strong IC conditions not only reduces the optimal profit under differentiated service but it may even make the optimal profit less than that under single service. In summary, we have the following key result from our analysis above: *even when customers have very high service valuations (so that capacity does not directly constrain the adequate provision of differentiated service), because of their self-selecting behavior, differentiated service can still perform worse than single service when they are not sufficiently heterogeneous.*

5. Duopoly

We consider two identical SPs (SP 1 and SP 2) with *price-plus-delay* competition between them. The customers select an SP (and a service class if the SP is offering differentiated service) based on

[6]The proof is similar to that in Proposition 4; we omit the details for the sake of conciseness.

their dis-utilities. For the purposes of focusing our analysis on the competitive dynamics, analytical tractability, and expositional clarity, we assume customer valuation v is high enough so that it does not constrain the price(s) charged by the SPs. However, note that the price(s) of an SP is/are still limited by the price(s) charged by the other SP. We denote the price charged by SP i $(i = 1, 2)$ under single service as p_i. The fraction of total customers who select single service from SP i and are of Type j $(j = 1, 2)$ is δ_{ij}; note that $0 \leq \delta_{i1} \leq q$ and $0 \leq \delta_{i2} \leq 1-q$. The prices for high-priority and low-priority services charged by SP i providing differentiated service are denoted by p_{ih} and p_{il} respectively. Then IC conditions similar to those in (9) and (10) have to be satisfied. Therefore, as in §4, only impatient (patient) customers select the high-priority (low-priority) service. The fraction of total customers who select high-priority and low-priority services from SP i are denoted by δ_{ih} and δ_{il} respectively; note that $0 \leq \delta_{ih} \leq q$ and $0 \leq \delta_{il} \leq 1-q$. Regardless of whether SP i provides single or differentiated service, we let δ_i denote the fraction of total customers that purchase from her. In the presence of two SPs, there are three possible ways in which service gets provided at *equilibrium* [7]: (i) both SPs provide single service, (ii) one of them provides single service while the other provides differentiated service, and (iii) both SPs provide differentiated service.

5.1 Single Service by Both SPs

SPs 1 and 2 charge prices p_1 and p_2 respectively. Then the profit maximization problem for SP 1 can be formulated as

$$
\begin{aligned}
(\mathcal{SS}) \qquad & \max_{\delta_{11},\delta_{12},p_1} \pi_{1,SS} = \lambda p_1 (\delta_{11} + \delta_{12}) \\
s.t.\ \delta_{11}\left(p_1 + \frac{\eta_1}{1-\lambda(\delta_{11}+\delta_{12})}\right) &\leq \delta_{11}\left(p_2 + \frac{\eta_1}{1-\lambda+\lambda(\delta_{11}+\delta_{12})}\right) \qquad (12)\\
\delta_{12}\left(p_1 + \frac{\eta_2}{1-\lambda(\delta_{11}+\delta_{12})}\right) &\leq \delta_{12}\left(p_2 + \frac{\eta_2}{1-\lambda+\lambda(\delta_{11}+\delta_{12})}\right) \qquad (13)\\
0 \leq \delta_{11} \leq q; 0 &\leq \delta_{12} \leq 1-q; p_1 \geq 0.
\end{aligned}
$$

Constraints (12) and (13) require respectively that if any impatient and patient customers buy from SP 1 then their dis-utilities should not exceed those obtained by purchasing from SP 2. Note that, because the valuation v is high, the market is covered and all the customers purchase from one of the SPs. First, we characterize the optimal values of the fractions in $\mathcal{SS}$ for any given values of $p_1, p_2 \geq 0$.

[7] The equilibrium here pertains to the SPs' decision of whether to provide single or differentiated service. Further, we note that (i) because this decision is strategic and usually a long-term one, we only consider *pure strategy* equilibrium and (ii) for any of SPs' service choice (and pricing) decisions, there may be a *sub-game equilibrium* among the self-selecting customers pertaining to their service choices.

Proposition 5 *Let* $\alpha_1 \equiv \arg\min_{0\le\alpha\le 1}\left|p_2 - p_1 + \eta_1\left(\frac{1}{1-\lambda+\lambda\alpha} - \frac{1}{1-\lambda\alpha}\right)\right|$ *and* $\alpha_2 \equiv \arg\min_{0\le\alpha\le 1}\left|p_2 - p_1 + \eta_2\left(\frac{1}{1-\lambda+\lambda\alpha} - \frac{1}{1-\lambda\alpha}\right)\right|$. *Optimal values of* δ_{11} *and* δ_{12} *for* $p_1, p_2 \ge 0$ *are then given by:*

1. *If* $p_1 + \frac{\eta_1}{1-\lambda} \le p_2 + \eta_1$ *then* $\delta_{11}^* = q$ *and* $\delta_{12}^* = 1 - q$.
2. *If* $p_1 < p_2$ *and* $p_1 + \frac{\eta_1}{1-\lambda} > p_2 + \eta_1$ *then* $\delta_{11}^* = (\alpha_1 - \delta_{12}^*)^+$ *and* $\delta_{12}^* = \min(\alpha_2, 1-q)$.
3. *If* $p_1 = p_2$ *then* $(\delta_{11}^*, \delta_{12}^*) = \{(\delta_{11}, \delta_{12}) : \delta_{11} + \delta_{12} = 0.5, 0 \le \delta_{11} \le q, 0 \le \delta_{12} \le 1 - q\}$.
4. *If* $p_1 > p_2$ *and* $p_1 + \eta_1 < p_2 + \frac{\eta_1}{1-\lambda}$ *then* $\delta_{11}^* = \min(\alpha_1, q)$ *and* $\delta_{12}^* = (\alpha_2 - \delta_{11}^*)^+$.
5. *If* $p_1 + \eta_1 \ge p_2 + \frac{\eta_1}{1-\lambda}$ *then* $\delta_{11}^* = \delta_{12}^* = 0$.

Further, if prices are p_1 *and* p_2 *then, after customers' self-selection, the fraction of total customers who buy from SP 1 and are of Type 1 and 2 are determined by* δ_{11}^* *and* δ_{12}^* *respectively.*

Proposition 5 shows that the optimal service delivery in $\mathcal{SS}$ has a structure which is very different from that in $\mathcal{S}$. In $\mathcal{S}$, patient customers always *get a preference* over impatient ones due to their lower delay sensitivity. However, in $\mathcal{SS}$, the relationship between prices p_1 and p_2 determines who gets preferred at SP 1. Proposition 5 implies that if $p_1 < p_2$ then patient customers get a preference and if $p_1 > p_2$ then impatient customers get a preference at SP 1. The rationale is as follows: *a low-priced SP serves more customers so that patient customers with a lower delay sensitivity get a preference while a high-priced SP satisfies less customers so that impatient customers with a higher delay sensitivity get a preference.* Proposition 5 also shows that although SP 1 cannot directly set the δ_{1j}'s, the corresponding fractions (at equilibrium) are given by δ_{1j}^*'s because of the self-selecting customer behavior. Also, note that except when $p_1 = p_2$, both δ_{11}^* and δ_{12}^* are *uniquely determined* by customers' self-selection. Next, we characterize the *higher level* equilibrium involving prices charged by SPs 1 and 2 when both of them provide single service.

Theorem 2 *If an equilibrium exists, it is unique and symmetric. Further, the outcome of price-plus-delay competition is characterized as follows: (i) if* $q \ge 0.5$ *then* $\tilde{p}_{1,SS} = \tilde{p}_{2,SS} = \frac{\eta_1\lambda}{\left(1-\frac{\lambda}{2}\right)^2}$ *and* $\tilde{\pi}_{1,SS} = \tilde{\pi}_{2,SS} = \frac{\eta_1\lambda^2}{2\left(1-\frac{\lambda}{2}\right)^2}$, *(ii) if* $q < 0.5$ *and* $\frac{\eta_2\lambda^2}{2\left(1-\frac{\lambda}{2}\right)^2} \ge \max_{0\le\delta_1\le q}\lambda\delta_1\left(\frac{\eta_2\lambda}{\left(1-\frac{\lambda}{2}\right)^2} + \frac{\eta_1}{1-\lambda+\lambda\delta_1} - \frac{\eta_1}{1-\lambda\delta_1}\right)$ *then* $\tilde{p}_{1,SS} = \tilde{p}_{2,SS} = \frac{\eta_2\lambda}{\left(1-\frac{\lambda}{2}\right)^2}$ *and* $\tilde{\pi}_{1,SS} = \tilde{\pi}_{2,SS} = \frac{\eta_2\lambda^2}{2\left(1-\frac{\lambda}{2}\right)^2}$, *and (iii) if* $q < 0.5$ *and* $\frac{\eta_2\lambda^2}{2\left(1-\frac{\lambda}{2}\right)^2} < \max_{0\le\delta_1\le q}\lambda\delta_1\left(\frac{\eta_2\lambda}{\left(1-\frac{\lambda}{2}\right)^2} + \frac{\eta_1}{1-\lambda+\lambda\delta_1} - \frac{\eta_1}{1-\lambda\delta_1}\right)$ *then there is no pure-strategy equilibrium.*

Theorem 2 shows that, depending on the composition of the customers, there are two possible types of equilibrium when SP 1 and SP 2 both provide single service. If there are many impatient

customers ($q \geq 0.5$) then there is a *high-price equilibrium*. The presence of a large fraction of such customers enables each SP to charge the *high price* because deviating from that price is going to be costly for the other SP. If there are few impatient customers (q is low) then there is a *low-price equilibrium*. The low value of q ensures that an SP would not benefit from charging a higher price and selling only to impatient customers when the other SP charges the *low price*. If q takes intermediate values then there is *no pure-strategy equilibrium* because the high price makes an SP better from reducing her price and increasing her sales while the low price makes her better from increasing her price and satisfying only impatient customers.

Next, we consider the price-plus-delay competition when both single and differentiated services are offered by the two SPs.

5.2 Single and Differentiated Services

We assume WLOG that SP 1 and SP 2 offer single and differentiated services respectively. Then the optimization problem for SP 1 can be formulated as

$$
\begin{aligned}
(\mathcal{SD}1) \qquad & \max_{\delta_{11},\delta_{12},p_1} \pi_{1,SD} = \lambda p_1 \left(\delta_{11} + \delta_{12}\right) \\
s.t.\ \delta_{11}\left(p_1 + \frac{\eta_1}{1-\lambda\left(\delta_{11}+\delta_{12}\right)}\right) &\leq \delta_{11}\left(p_{2h} + \frac{\eta_1}{1-\lambda q+\lambda\delta_{11}}\right) \qquad (14)\\
\delta_{12}\left(p_1 + \frac{\eta_2}{1-\lambda\left(\delta_{11}+\delta_{12}\right)}\right) &\leq \delta_{12}\left(p_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\delta_{11}\right)\left(1-\lambda+\lambda(\delta_{11}+\delta_{12})\right)}\right) \qquad (15)\\
0 \leq \delta_{11} \leq q; 0 &\leq \delta_{12} \leq 1-q; p_1 \geq 0.
\end{aligned}
$$

Constraints (14) and (15) respectively require that patient and impatient customers, if any of them select SP 1, incur a lower dis-utility. We refer to constraint (14) (constraint (15)) as *non-trivial* if $\delta_{11} > 0$ ($\delta_{12} > 0$). Next, we characterize the optimal values of δ_{11} and δ_{12} for any given prices.

Proposition 6 *Let* $\beta_1 \equiv \arg\min_{0\leq\beta\leq 1}\left|p_{2h} - p_1 + \eta_1\left(\frac{1}{1-\lambda q+\lambda\beta} - \frac{1}{1-\lambda\beta}\right)\right|$ *and* $\beta_2 \equiv \arg\min_{0\leq\beta\leq 1}\left|p_{2l} - p_1 + \eta_2\left(\frac{1}{(1-\lambda q)(1-\lambda+\lambda\beta)} - \frac{1}{1-\lambda\beta}\right)\right|$. *Optimal values of* δ_{11} *and* δ_{12} *for* $p_1, p_{2h}, p_{2l} \geq 0$ *are given by: (i) if* $\beta_1 \leq \beta_2$ *then* $\delta_{11}^* = (\beta_1 - \delta_{12}^*)^+$ *and* $\delta_{12}^* = \min(\beta_2, 1-q)$ *and (ii) if* $\beta_1 > \beta_2$ *then* $\delta_{11}^* = \min(\beta_1, q)$ *and* $\delta_{12}^* = (\beta_2 - \delta_{11}^*)^+$. *There are four possibilities: (i)* $\delta_{11}^* = 0$, *(ii)* $\delta_{11}^* = q$, *(iii)* $\delta_{12}^* = 0$, *or (iv)* $\delta_{12}^* = 0$. *Further, for any given prices, the fractions of impatient and patient customers satisfied by SP 1 get uniquely determined as* δ_{11}^* *and* δ_{12}^*.

Proposition 6 shows that at least one customer type, impatient or patient, exhibits an *all-or-none* strategy in selecting SP 1 who provides single service. *Either all the customers from this customer type select SP 1 or none of them do so.* Importantly, we note that this strategy is not a result of

any rule enforced by the SP but it is actually a consequence of self-selection by customers. The optimization problem for SP 2 can be formulated as

$$
\begin{aligned}
(\mathcal{SD}2) \qquad & \max_{\delta_{2h},\delta_{2l},p_{2h},p_{2l}} \pi_{2,SD} = \lambda\left(p_{2h}\delta_{2h} + p_{2l}\delta_{2l}\right) \\
s.t.\ p_{2h} + \frac{\eta_1}{1-\lambda\delta_{2h}} &\leq p_1 + \frac{\eta_1}{1-\lambda\left(1-\delta_{2h}-\delta_{2l}\right)} && (16)\\
p_{2l} + \frac{\eta_2}{\left(1-\lambda\delta_{2h}\right)\left(1-\lambda\delta_{2h}-\lambda\delta_{2l}\right)} &\leq p_1 + \frac{\eta_2}{1-\lambda\left(1-\delta_{2h}-\delta_{2l}\right)} && (17)\\
p_{2h} - p_{2l} &\leq \eta_1\left(\frac{\lambda(\delta_{2h}+\delta_{2l})}{1-\lambda\delta_{2h}-\lambda\delta_{2l}}\right) && (18)\\
p_{2h} - p_{2l} &\geq \eta_2\left(\frac{\lambda(\delta_{2h}+\delta_{2l})}{\left(1-\lambda\delta_{2h}-\lambda\delta_{2l}\right)^2}\right) && (19)\\
0 \leq \delta_{2h} \leq q; 0 &\leq \delta_{2l} \leq 1-q; p_{2h}, p_{2l} \geq 0.
\end{aligned}
$$

Constraints (16) and (16) ensure that impatient and patient customers, who select high-priority and low-priority service respectively, incur less dis-utility than they would have obtained from SP 1. Constraints (18) and (19) are the *strong* incentive compatibility constraints. Next, we identify some key properties of the equilibrium resulting from the price-plus-delay competition involving SPs 1 and 2.

Theorem 3 *At equilibrium, if it exists, SP 1 serves only one customer type and SP 2 satisfies either all the patient customers or all the impatient customers. Suppose* $\eta_2/\eta_1 < 1-\lambda$.
1. Let $\mathcal{F}_1(q) \equiv \frac{\eta_1\lambda q}{2-\lambda q} + \frac{\eta_2}{1-\lambda+\lambda q} - \frac{2\eta_2\lambda(2-q)}{(2-2\lambda+\lambda q)^2} - \frac{\eta_2}{(1-\lambda q)^2}$ *and* $\mathcal{F}_2(q) \equiv (1-\lambda q)^3 - \lambda q(1-\lambda+\lambda q)^2$. *Offering differentiated service is better for SP 1 than serving impatient customers if* $\mathcal{F}_1(q), \mathcal{F}_2(q) > 0$.
2. Let $\mathcal{F}_3(q) \equiv \frac{\eta_1\lambda q}{1-\lambda q} + \eta_2 + \frac{\eta_2\lambda q}{(1-\lambda q)^2} - \frac{\eta_2(2+\lambda-3\lambda q)}{(1-\lambda q)^2(2-\lambda-\lambda q)}$ *and* $\mathcal{F}_4(q) \equiv \frac{2\eta_1\lambda q}{1-\lambda q} + \frac{2\eta_2}{2-\lambda+\lambda q} - \frac{2\eta_2}{(1-\lambda q)(2-\lambda-\lambda q)} - \frac{\eta_2\lambda(1-q)}{(1-\lambda+\lambda q)^2}$. *Offering differentiated service is better for SP 1 than serving patient customers if* $\mathcal{F}_3(q), \mathcal{F}_4(q) > 0$.

Theorem 3 derives sufficient conditions so that offering differentiated service, in comparison to an equilibrium outcome, is beneficial for SP 1. These conditions impose some restrictions on the amount of heterogeneity among customers, arrival rate λ (which is the same as utilization since capacity is one), and the composition of (patient vs. impatient) customers. In particular, we note that if there is very high heterogeneity ($\eta_2 << \eta_1$) then SP 1 is better from offering differentiated service provided the fraction of impatient customers q is not high so that $\mathcal{F}_2(q) > 0$.

Figure 5 illustrates an example which demonstrates how $\mathcal{F}_i$'s ($i = 1, ..., 4$) vary with q. We find that $\mathcal{F}_1$, $\mathcal{F}_3$, and $\mathcal{F}_4$ are increasing in q, which yields *lower thresholds* for q above which all the $\mathcal{F}_i$'s become positive. However, $\mathcal{F}_2$ is decreasing in q, which yields an *upper threshold* for q below which

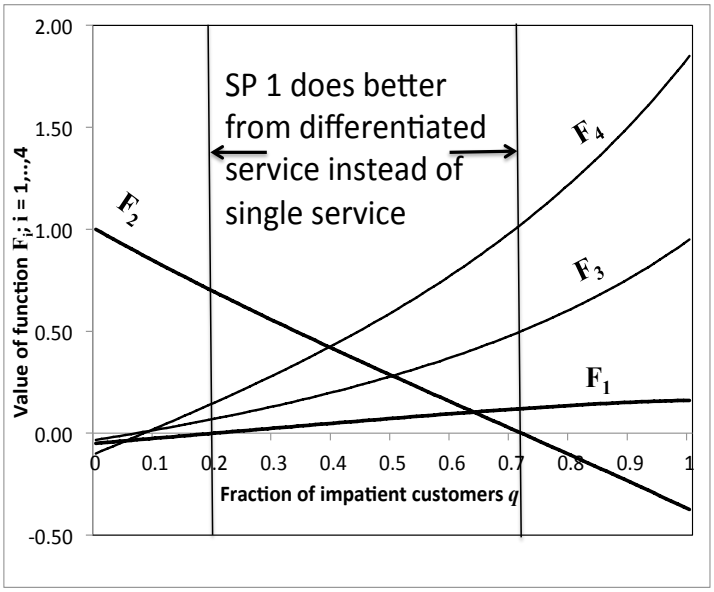

Figure 5: Variation of $\mathcal{F}_i$ $(i = 1, ..., 4)$ with q when $\lambda = 0.5$, $\eta_1 = 1$, and $\eta_2 = 0.05$

$\mathcal{F}_2$ becomes positive. All the $\mathcal{F}_i$'s become positive when q takes intermediate values in between 0.2 and 0.72 (correct to two decimals). The reason why differentiated service dominates single service (for SP 1 when SP 2 is offering differentiated service) for intermediate values of q is as follows: *if q is low or high then there are many customers of a single type, and hence, offering single service to those customers, instead of providing differentiated service, may be beneficial for SP 1.*

5.3 Differentiated Service by Both SPs

Both SPs provide differentiated service. SP 1 (SP 2) charges prices p_{1h} (p_{2h}) and p_{1l} (p_{2l}) for high-priority and low-priority services respectively. We assume that $\eta_1/\eta_2 \leq 1-\lambda$ because (i) it simplifies the analysis and (ii) even under a monopoly, when customers have high valuation, $\eta_1/\eta_2 > 1-\lambda$ results in differentiated service being sub-optimal (see §4.3). The optimization problem for SP 1[8] can be formulated as

$$
\begin{aligned}
(\mathcal{DD}1) \qquad & \max_{\delta_{1h},\delta_{1l},p_{1h},p_{1l}} \pi_{1,DD} = \lambda\left(p_{1h}\delta_{1h} + p_{1l}\delta_{1l}\right) \\
s.t.\ p_{1h} + \frac{\eta_1}{1-\lambda\delta_{1h}} &\leq p_{2h} + \frac{\eta_1}{1-\lambda q+\lambda\delta_{1h}} && (20)\\
p_{1l} + \frac{\eta_2}{\left(1-\lambda\delta_{1h}\right)\left(1-\lambda\delta_{1h}-\lambda\delta_{1l}\right)} &\leq p_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\delta_{1h}\right)\left(1-\lambda+\lambda\delta_{1h}+\lambda\delta_{1l}\right)} && (21)\\
p_{1h} - p_{1l} &\leq \eta_1\left(\frac{\lambda(\delta_{1h}+\delta_{1l})}{1-\lambda\delta_{1h}-\lambda\delta_{1l}}\right) && (22)\\
p_{1h} - p_{1l} &\geq \eta_2\left(\frac{\lambda(\delta_{1h}+\delta_{1l})}{\left(1-\lambda\delta_{1h}-\lambda\delta_{1l}\right)^2}\right) && (23)\\
0 \leq \delta_{1h} \leq q; 0 &\leq \delta_{1l} \leq 1-q; p_{1h}, p_{1l} \geq 0.
\end{aligned}
$$

[8]The formulation for SP 2 is similar and we omit it for the sake of conciseness.

Since SP 1 and SP 2 both provide differentiated service, we only consider cases in which patient as well as impatient customers select both SPs (the former select low-priority service while the latter choose high-priority service)[9], and hence $\exists \delta_{1h}, \delta_{1l}$, which are unique, with $0 < \delta_{1h} < q$ and $0 < \delta_{1l} < 1 - q$ such that $p_{1h} + \frac{\eta_1}{1-\lambda\delta_{1h}} = p_{2h} + \frac{\eta_1}{1-\lambda q+\lambda\delta_{1h}}$ and $p_{1l} + \frac{\eta_2}{(1-\lambda\delta_{1h})(1-\lambda\delta_{1h}-\lambda\delta_{1l})} = p_{2l} + \frac{\eta_2}{(1-\lambda q+\lambda\delta_{1h})(1-\lambda+\lambda\delta_{1h}+\lambda\delta_{1l})}$. Note that if the firms set these prices and the corresponding IC constraints are satisfied then the unique δ_{ij}'s result from the self-selection of customers. Next, we establish a key property regarding the profit function of SP 1.

Proposition 7 *Let* $\mathcal{G}(x,y) \equiv \lambda x(p_{2h} + \frac{\eta_1}{1-\lambda q+\lambda x} - \frac{\eta_1}{1-\lambda x}) + \lambda(y-x)(p_{2l} + \frac{\eta_2}{(1-\lambda q+\lambda x)(1-\lambda+\lambda y)} - \frac{\eta_1}{(1-\lambda x)(1-\lambda y)})$. *Then* $\mathcal{G}$ *is strictly concave in* $x\ \forall 0 \le y \le 1$ *and it is also strictly concave in* $y\ \forall 0 \le x \le 1$. *If* $(x^*, y^*) = \arg\max_{0\le x,y\le 1} \mathcal{G}(x,y)$ *then* x^* *and* y^* *satisfy:*

$$y^* = \mathcal{H}_1(x^*) = \frac{1}{\lambda}\left(\lambda - 1 + \frac{\lambda(1-q)}{\frac{\eta_1}{\eta_2}(1-\lambda q) - 1 + (1-\lambda q+\lambda x^*)^2\left(\frac{p_{2h}-p_{2l}}{\eta_2} - \frac{\eta_1-\eta_2}{\eta_2(1-\lambda x^*)^2}\right)}\right) \quad (24)$$

$$x^* = \mathcal{H}_2(y^*) = \frac{1}{\lambda}\left(\lambda q - 1 + \frac{\lambda(1-q)}{1 + (1-\lambda+\lambda y^*)^2\left(\frac{p_{2l}}{\eta_2} - \frac{1}{(1-\lambda y^*)^2}\right)}\right) \quad (25)$$

If $\mathcal{H}_2(\mathcal{H}_1(0)) < 0$, *then there exists at most a single* (x^*, y^*). *This condition is satisfied when*

$$\frac{\eta_2}{\eta_1} \le \frac{q\,(1-\lambda q)\,(2-\lambda)\,\sqrt{1-\lambda}}{q\,(2-\lambda)\,(2-\lambda q)\,\sqrt{1-\lambda} + (1-q)\left(\sqrt{1-\lambda q} + \sqrt{1-\lambda}\right)}. \quad (26)$$

The profit of SP 1, $\pi_{1,DD}$, is obtained by setting $x = \delta_{1h}$ and $y = \delta_1 = \delta_{1h} + \delta_{1l}$. Proposition 7 establishes sufficient conditions under which SP 1 has unique "*best response prices*" for any prices charged by SP 2[10]. In particular, we observe that if the customers are highly heterogeneous so that η_1/η_2 is high then SP 1 has a unique best response function. Next, we characterize the equilibrium resulting from price-plus-delay competition between SP 1 and SP 2 when both of them provide differentiated service.

Theorem 4 *The prices* $\tilde{p}_{1h} = \tilde{p}_{2h} = \frac{4\eta_1\lambda q}{(2-\lambda q)^2} + \frac{8\eta_2\lambda(1-q)(4-\lambda-\lambda q)}{(2-\lambda q)^2(2-\lambda)^2}$ *and* $\tilde{p}_{1l} = \tilde{p}_{2l} = \frac{8\eta_2\lambda(1-q)}{(2-\lambda q)(2-\lambda)^2}$ *result in a unique symmetric equilibrium with* $\tilde{\delta}_{1h} = \tilde{\delta}_{2h} = q/2$ *and* $\tilde{\delta}_{1l} = \tilde{\delta}_{2l} = (1-q)/2$ *if* $\frac{2\eta_2\lambda}{(2-\lambda)^2} \le \tilde{p}_{1h} - \tilde{p}_{1l} \le \frac{\eta_1\lambda}{(2-\lambda)}$ *and* $\mathcal{H}_2(\mathcal{H}_1(0))|_{p_{2h}=\tilde{p}_{2h}, p_{2l}=\tilde{p}_{2l}} < 0$.

5.4 Service Delivery under Duopoly: Single Service vs. Single and Differentiated Services vs. Differentiated Service

The SPs can choose to provide either single or differentiated service. This choice precedes their pricing decisions, which in turn precedes customers' choices. Therefore, we have a three-stage game

[9] If either patient or impatient customers do not choose an SP then she is better by offering single service instead of differentiated service.

[10] Note that the fractions δ_{1h} and δ_{1l} get uniuqely determined from customers' self-selection for any given prices.

SP 1/SP 2	Single Service	Differentiated Service
Single Service	$\tilde{\pi}_1 = \tilde{\pi}_2 = 0.2222$ $\tilde{p}_1 = \tilde{p}_2 = 0.8889$ $\tilde{\delta}_1 = \tilde{\delta}_2 = 0.5$	$\tilde{\pi}_1 = 0.0421$, $\tilde{\pi}_2 = 0.1227$ $\tilde{p}_1 = 0.3318$, $\tilde{p}_{2h} = 0.337$, $\tilde{p}_{2l} = 0.3251$ $\tilde{\delta}_{11} = 0.254$, $\tilde{\delta}_{12} = 0$, $\tilde{\delta}_{2h} = 0.246$, $\tilde{\delta}_{2l} = 0.5$
Differentiated Service	$\tilde{\pi}_1 = 0.1227$, $\tilde{\pi}_2 = 0.0421$ $\tilde{p}_{1h} = 0.337$, $\tilde{p}_{1l} = 0.3251$, $\tilde{p}_2 = 0.3318$ $\tilde{\delta}_{1h} = 0.246$, $\tilde{\delta}_{1l} = 0.5$, $\tilde{\delta}_{21} = 0.254$, $\tilde{\delta}_{22} = 0$	$\tilde{\pi}_1 = \tilde{\pi}_2 = 0.0426$ $\tilde{p}_{1h} = \tilde{p}_{2h} = 0.336$, $\tilde{p}_{1l} = \tilde{p}_{2l} = 0.0051$ $\tilde{\delta}_{1h} = \tilde{\delta}_{2h} = \tilde{\delta}_{1l} = \tilde{\delta}_{2l} = 0.25$

Table 1: Equilibrium profits, prices, and fractions of customers under different types of service delivery when $\eta_1 = 1$, $\eta_2 = 0.01$, and $\lambda = q = 0.5$

and the SPs make the first stage service delivery decisions based on the equilibriums resulting from price-plus-delay competition between them and customer choices. There are four possible sub-games from the service delivery decisions, which were analyzed in §5.1-5.3[11]. Next, we present an example to understand the trade-offs in and gain insights from the equilibrium involving SPs' service delivery decisions.

We consider an example in which $\eta_1 = 1$, $\eta_2 = 0.01$, and $\lambda = q = 0.5$. Note that $\eta_2/\eta_1 << 1-\lambda$ and there is *high heterogeneity* between patient and impatient customers. Table 1 characterizes the sub-game equilibriums under different types of service delivery. When both SPs offer single service, because the fraction of impatient customers is high ($q \geq 0.5$), there is a *high-price equilibrium*. When one of them offers single service and the other SP offers differentiated service, we numerically find that there is a unique equilibrium. In this equilibrium, the single service SP only satisfies some (but not all) impatient customers (interestingly, we find that there is no equilibrium in which the single service SP satisfies some (but not all) patient customers). From Table 1, we also find that *the profits of both SPs are significantly lower than those obtained when they both provide single service.*

The result that an SP providing single service loses from the other one differentiating its service (through prioritization) has an intuitive explanation. However, it is somewhat counterintuitive that even the SP providing differentiated service loses, and it is explained as follows. Because the SP providing single service satisfies only impatient customers, it prices much more aggressively than under the high-price equilibrium characterized above. This aggressive pricing intensifies the competition between the SPs so much that in spite of selling to more customers (e.g., $\tilde{\delta}_2 = 0.746 > 0.5$ when SP 1 and SP 2 provide single and differentiated services respectively), the SP providing differentiated service obtains less profit. The effect of intensified competition (from provision of differentiated service) is further illustrated by comparing the profits and sales when an SP offers

[11]The cases **SD** and **DS** result in symmetric equilibrium outcomes.

single and differentiated services, and the other SP provides differentiated service. From Table 1, we find that the increase in profit is marginal (e.g., when SP 2 offers differentiated service, $\tilde{\pi}_1$ changes from 0.0421 to 0.0426 when SP 1's service delivery changes from single service to differentiated service) even though the sales increase significantly (correspondingly, $\tilde{\delta}_1$ increases from 0.254 to 0.5). The sales increase is from providing low-priority service to patient customers but they are charged a much lower price ($\tilde{p}_{1l} = 0.0051 << 0.336 = \tilde{p}_{1h}$) due to more intense competition for them between the SPs, thereby resulting in just a marginal increase in profit.

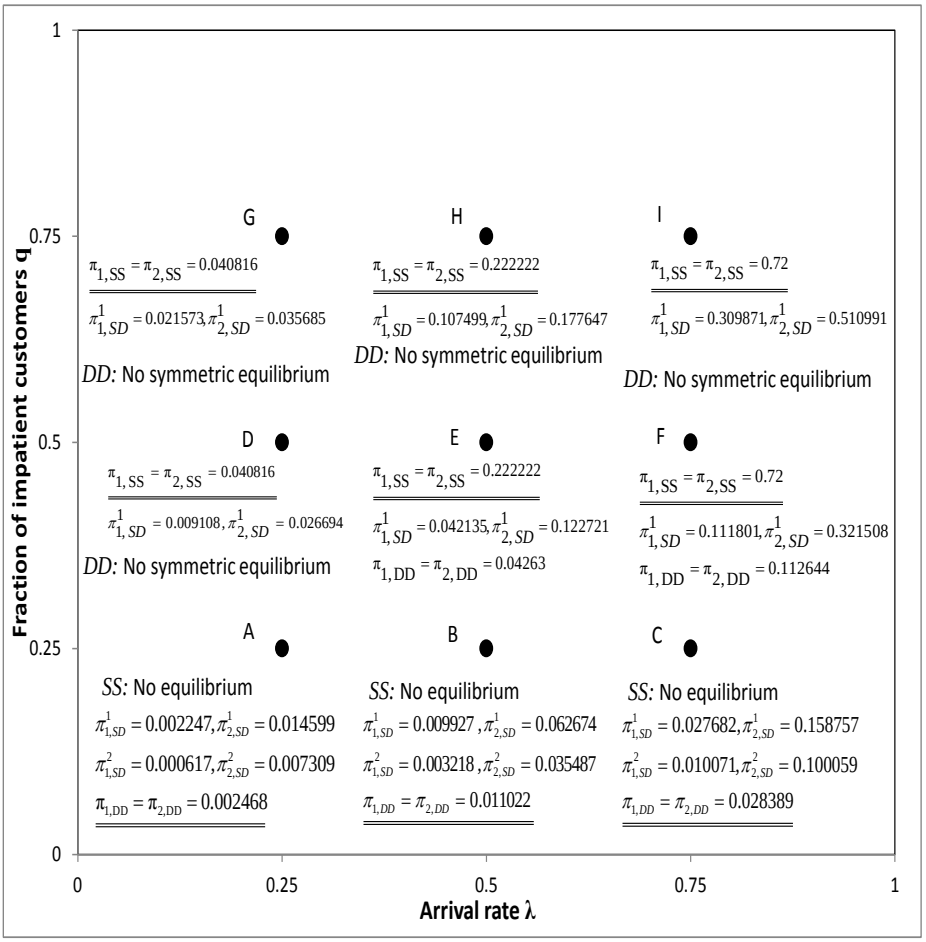

Figure 6: Equilibrium profits under different service deliveries, and how they change with λ and q; superscripts ^{1} and ^{2} denote possible equilibriums in which the SP providing single service satisfies only impatient and patient customers respectively, the overall equilibrium is marked with $=$

Table 1 shows that providing single service *dominates* providing differentiated service for both SPs. Hence both of them provide single service at equilibrium *even though* $\eta_2/\eta_1 << 1-\lambda$ *and there is high heterogeneity between patient and impatient customers.* Provision of differentiated service through prioritization of customers has two effects: (i) it enables an SP to better differentiate her customers and (ii) it intensifies the competition between the SPs. In this example, we find that the

loss from the second effect outweighs any gain from the first effect, thereby making differentiated service sub-optimal for both SPs.

6. Discussion

We provide more examples to analyze the equilibrium involving service delivery decisions and how the parameters λ and q, the arrival rate of customers and the fraction of impatient customers respectively, affect them. In all these examples, $\eta_1 = 1$ and $\eta_2 = 0.01$ so that there is high heterogeneity between patient and impatient customers. Figure 6 summarizes the results. In examples A through C in which $q = 0.25$, we find that there is no equilibrium when both SPs provide single service because q is too high for a low-price equilibrium but it is less than 0.5 so that there is no high-price equilibrium. The equilibrium service delivery[12] then is for both SPs to provide differentiated service. In these examples, unlike the analysis in §5.4, we find that the positive effect on profit from differentiating customers outweighs the negative effect from more intense competition. That can be observed from examining the equilibrium when the SPs provide single and differentiated services. From Figure 6, we find that there are two possibilities: (i) the single service SP satisfies only impatient customers and (ii) she satisfies only patient customers (indicated by ^{1} and ^{2} respectively in Figure 6). We also find that, since η_2/η_1 is low, the profits of both SPs are higher when the single service SP satisfies only impatient customers (e.g., in example B $\pi^1_{1,SD} = 0.009927 > \pi^2_{1,SD} = 0.003218$ and $\pi^1_{2,SD} = 0.062674 > \pi^2_{2,SD} = 0.035487$) so that it becomes the equilibrium. However, the single service SP is still better from providing differentiated service (in example B $\pi^1_{1,SD} = 0.009927 < \pi_{1,DD} = 0.011022$). In contradiction, examples D through I are all akin to the example in §6 (note that example E is the same as the one in §6), and we find that the loss from intense competition outweighs the benefit from differentiated service. In conclusion, we note that although all the examples in Figure 6 result in at least one SP being better under equilibrium service delivery, a *prisoner's dilemma* scenario is also possible. For instance, when $\lambda = 0.75$ and $q = 0.4$, there is no equilibrium under **SS**; $\pi^1_{1,SD} = 0.0779$, $\pi^2_{1,SD} = 0.3149$ (there is no equilibrium when SP 1 only satisfies patient customers); and $\pi_{1,DD} = \pi_{2,DD} = 0.0689$. Although $\pi^1_{i,SD} > \pi_{i,DD}$, **SD** is not an equilibrium because we find that $\mathcal{F}_j(q) > 0; j = 1, ..., 4$ and hence Theorem 3 implies that SP 1 is better by offering differentiated service, thereby resulting in **DD** as the equilibrium service delivery.

[12] We consider **SS** only if it has a price equilibrium and we consider **DD** only if it has a symmetric price equilibrium.

7. Conclusion

Time is used as a differentiating factor by service providers (SPs) under different contexts. We consider differentiated service with two service classes in which some customers are prioritized, and analyze when offering it in the presence of self-selecting customers, who maximize their individual expected utilities, would be beneficial for SPs. The customers are heterogeneous in their delay sensitivities and belong to one of two types: impatient and patient. We first show that the prices for high-priority and low-priority services have to satisfy *strong* incentive compatibility (IC) conditions; otherwise, the queuing dynamics and customers' self-selection result in all of them selecting a single service class (high-priority or low-priority). We also show that these conditions are stronger than those IC conditions which *preclude customers from changing their service classes after they have made their decisions*. In the presence of these conditions, we examine how differentiated service performs vis-a-vis single service under both monopoly and duopoly.

When a single SP sells to customers, we characterize the SP's optimal pricing decisions under both single and differentiated services. We find that self-selection results in patient customers *getting a preference* over impatient ones under single service. Under differentiated service, we observe that both the amount of customers selecting high-priority service and the total amount of customers getting satisfied increase as customers' valuation increases; however, the amount of customers selecting low-priority service can decrease. In comparing single and differentiated services, we find that even if the customers' valuation is very high, single service still outperforms differentiated service when the customers are *not sufficiently heterogeneous*. Unlike prior research on queuing optimization in which prioritization is either always better or is not possible due to limited capacity, we establish a new criterion for it to perform better in the presence of self-selecting customers: *sufficient customer heterogeneity*[13].

When two SPs sell to customers, they engage in price-plus-delay competition and three types of service delivery are possible: both of them provide single service (**SS**), one of them provides single service and the other provides differentiated service (**SD** and **DS**), and both of them provide differentiated service (**DD**). We identify two kinds of equilibrium under **SS**: *low-price* and *high-price*. We also characterize the symmetric equilibrium under **DD**. We derive sufficient conditions for which the single service SP in SD or DS would be better providing differentiated service (in comparison to the equilibrium under **SD** or **SD**) and hence **SD** or **DS** would be dominated by **DD**. In comparing the three types of service delivery, we find that the equilibrium service delivery is

[13]Customers cannot be just heterogeneous, they have to be sufficiently heterogeneous so that $\eta_2/\eta_1 \leq 1-\lambda$.

generally either **SS** or **DD**. Further, we also find that, due to the competitive dynamics between the SPs, sufficient customer heterogeneity is no longer enough, and the performance of differentiated service vis-a-vis single service also depends on the fraction of impatient customers q. When an SP provides differentiated service, it can increase her profit from better customer differentiation but it can also decrease her profit from intensified competition with the other SP. We show that, generally, when q is low **DD** is the equilibrium service delivery and the SPs also benefit from differentiating customers but when q is high SS is the equilibrium service delivery and the SPs benefit from the high-price equilibrium.

References

Afeche, P. 2013. Incentive-compatible revenue management in queueing systems: optimal strategic delay. *Manufacturing and Service Operations Management* **15**(3) 423–443.

Afeche, P., O. Baron, Y. Kerner. 2013. Pricing time-sensitive services based on realized performance. *Manufacturing and Service Operations Management* **15**(3) 492–506.

Allon, G., A. Federgruen. 2009. Competition in service industries with segmented markets. *Management Science* **55**(4) 619–634.

Anand, K. S., M. F. Pac, S. Veeraraghavan. 2011. Quality-speed conundrum: trade-offs in customer-intensive services. *Management Science* **57**(1) 40–56.

Armony, M., M. Haviv. 2003. Price and delay competition between two service providers. *European Journal of Operational Research* **147**(1) 32–50.

Boyaci, T., S. Ray. 2003. Product differentiation and capacity cost interaction in time and price sensitive markets. *Manufacturing and Service Operations Management* **5**(1) 18–36.

Chen, H., Y. W. Wan. 2003. Price competition of make-to-order firms. *IIE Transactions* **35**(9) 817–832.

Gavirneni, S., V. G. Kulkarni. 2014. Self-selecting priority queues with burr distributed waiting costs. Http://papers.ssrn.com/sol3/papers.cfm?abstract _ id=2499146.

Hassin, R., M. Haviv. 2003. *To queue or not to queue: Equilibrium behavior in queueing systems*, vol. 59. Kluwer Academic Publishers.

Hsu, V. N., S. H. Xu, B. Jukic. 1998. Optimal scheduling and incentive compatible pricing for a service system with quality of service guarantees. *Manufacturing and Service Operations Management* **11**(3) 375–396.

Katta, A., J. Sethuraman. 2005. Incentive-compatible revenue management in queueing systems: optimal strategic delay. Tech. Rep. TR-2005-04, CORC, Columbia University.

Knowledge@Wharton, . 2013. Skipped out on your restaurant reservation? That will be $200, please. Http://business.time.com/2013/06/08/skipped-out-on-your-restaurant-reservation-that-will-be-200-please/.

Lederer, P. J., L. Li. 1997. Pricing, production, scheduling, and delivery-time competition. *Operations Research* **45**(3) 407–420.

Li, L., L. Jiang, L. Liu. 2012. Service and price competition when customers are naive. *Production and Operations Management* **21**(4) 747–760.

Li, L., Y. S. Lee. 1994. Pricing and delivery-time performance in a competitive environment. *Management Science* **40**(3) 633–646.

Luski, I. 1976. On partial equilibrium in a queuing system with two servers. *The Review of Economic Studies* **43**(3) 519–525.

Markovich, M. 2014. Pay up and show up: Local restaurant charges for reservations. Http://www.komonews.com/news/local/Pay-up-and-show-up-Seattle-restaurant-charges-for-making-reservations-274556091.html.

Pangburn, M. S., E. Stavrulaki. 2008. Capacity and price setting for dispersed, time-sensitive customer segments. *European Journal of Operational Research* **184**(3) 1100–1121.

Rao, S., E. R. Peterson. 1998. Optimal pricing of priority services. *Operations Research* **46**(1) 46–56.

Sainathan, A. 2014. Customer differentiation in ancillary services? free service, prioritization, and strategic delay. Http://papers.ssrn.com/sol3/papers.cfm?abstract_ id=2494681.

So, K. C., J. S. Song. 1998. Price, delivery time guarantees, and capacity selection. *European Journal of Operational Research* **111**(1) 28–49.

Wallop, H. 2010. £350 to queue jump at a theme park. Http://www.telegraph.co.uk/finance/newsbysector/retailandconsumer/7821388/350-to-queue-jump-at-a-theme-park.html.

Wieczner, J. 2013. Pros and cons of concierge medicine. Http://www.wsj.com/articles/SB10001424052702303471004579165470633112630.

Zhang, Z., D. Dey, Y. Tan. 2007. Optimal scheduling and incentive compatible pricing for a service system with quality of service guarantees. *INFORMS Journal on Computing* **19**(2) 248–260.

Appendix

Proof (Proof of Proposition 1) Under $M/M/1$ preemptive priority, $W_l(x, y-x) - W_h(x) = 1/((1-\lambda x)\cdot(1-\lambda y)) - 1/(1-\lambda x) = \lambda y/((1-\lambda x)\cdot(1-\lambda y))$, which is strictly increasing in $x\ \forall y > 0$. Similarly, under $M/M/1$ non-preemptive priority, $W_l(x, y-x) - W_h(x) = (\lambda y/((1-\lambda x)\cdot(1-\lambda y)) + 1) - (\lambda y/(1-\lambda x) + 1) = (\lambda y)^2/((1-\lambda x)\cdot(1-\lambda y))$, which is strictly increasing in $x\ \forall y > 0$. Suppose $0 \le x_1 < x_2 \le y$. Then $W_l(x_1, y-x_1) - W_h(x_1) < W_l(x_2, y-x_2) - W_h(x_2)$. Using $x_1 = 0$, $x_2 = \delta_h$, $y = \delta_h + \delta_l$, and the fact that $W_l(0, x) = W(x)$, we get $W(\delta_h + \delta_l) - W_h(0) \le W_l(\delta_h, \delta_l) - W_h(\delta_h)$. So the RHS of (1) does not exceed the RHS of (3) and hence (1) $\Rightarrow$ (3). Similarly, using $x_1 = \delta_h$, $x_2 = \delta_h + \delta_l$, $y = \delta_h + \delta_l$, and the fact that $W_h(\delta_h + \delta_l) = W(\delta_h + \delta_l)$, we get $W_l(\delta_h, \delta_l) - W_h(\delta_h) \le W_l(\delta_h + \delta_l, 0) - W(\delta_h + \delta_l)$. So the RHS of (2) is greater than or equal to the RHS of (4) and hence (2) $\Rightarrow$ (4). ∎

Proof (Proof of Proposition 2) The result follows from the fact that $W_l(x, y-x) - W_h(x)$ is strictly increasing in $x\ \forall x \in [0, y], y > 0$ (see Proposition 1). We prove it by contradiction. Suppose some impatient customers select low-priority service. Then they should obtain a higher net-utility from their choice so that $p_l + \eta_1 W_l(\delta_h, \delta_l) \le p_h + \eta_1 W_h(\delta_h)$, i.e., $p_h - p_l \ge \eta_1 (W_l(\delta_h, \delta_l) - W_h(\delta_h))$. However, that condition violates (1) because $W_l(\delta_h, \delta_l) - W_h(\delta_h) > W_l(0, \delta_h + \delta_l) - W_h(0) = W(\delta_h + \delta_l) - W_h(0)\ \forall \delta_h > 0$. Similarly, if some patient customers select high-priority service then we require that $p_h + \eta_2 W_h(\delta_h) \le p_l + \eta_2 W_l(\delta_h, \delta_l)$, i.e., $p_h - p_l \le \eta_2 (W_l(\delta_h, \delta_l) - W_h(\delta_h))$. However, it violates (2) because $W_l(\delta_h, \delta_l) - W_h(\delta_h) < W_l(\delta_h + \delta_l, 0) - W_h(\delta_h + \delta_l) = W_l(\delta_h + \delta_l, 0) - W(\delta_h + \delta_l)\ \forall \delta_l > 0$. ∎

Proof (Proof of Proposition 3) If $\delta \le 1 - q$ then (6) binds under optimality because otherwise price p can be increased yielding better profit. The profit then becomes $\lambda\delta\left(v - \frac{\eta_2}{1-\lambda\delta}\right)$ which is strictly concave in δ. Similarly if $\delta > 1 - q$ then (5) binds (constraint (5) is stronger than constraint (6)) and the profit is $\lambda\delta\left(v - \frac{\eta_2}{1-\lambda\delta}\right)$ which is again strictly concave in δ. So the profit is piecewise concave (with two pieces) and it is discontinuous at $\delta = 1 - q$. The result in Proposition 3 is then obtained by comparing the profits at the appropriate interior solution(s)/boundary solution(s). ∎

Proof (Proof of Proposition 4) Under optimality, there are three possibilities: (i) constraints (7) and (10) bind, (ii) (7) and (8) bind, or (iii) (7) and (10) bind. Consider the first case with $p_h = v - \frac{\eta_1}{1-\lambda\delta_h}$ and $p_l = v - \frac{\eta_1}{1-\lambda\delta_h} - \eta_2\left(\frac{\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h-\lambda\delta_l)^2}\right)$. We show that this solution is feasible. Clearly, it satisfies (9), and hence,

we just need to show that (8) holds. It holds because

$$\begin{aligned} p_l + \frac{\eta_2}{(1-\lambda\delta_h)\cdot(1-\lambda\delta_h-\lambda\delta_l)} &= v - \frac{\eta_1}{1-\lambda\delta_h} - \frac{\eta_2\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h-\lambda\delta_l)^2} + \frac{\eta_2}{(1-\lambda\delta_h)\cdot(1-\lambda\delta_h-\lambda\delta_l)} \\ &= v - \frac{\eta_1}{1-\lambda\delta_h} + \frac{\eta_2\left(1-\lambda(\delta_h+\delta_l)(2-\lambda\delta_h)\right)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)^2} \\ &< v - \frac{\eta_2}{1-\lambda\delta_h} + \frac{\eta_2\left(1-\lambda(\delta_h+\delta_l)(2-\lambda\delta_h)\right)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)^2} \\ &= v - \frac{\eta_2\lambda\delta_l(\lambda\delta_h+\lambda\delta_l)}{(1-\lambda\delta_h)(1-\lambda\delta_h-\lambda\delta_l)^2} < v. \end{aligned}$$

Similarly, it can be shown that the second and third cases lead to infeasible solutions. When(7) and (10) bind, the profit becomes $\pi_D = \lambda\delta_h\left(v - \frac{\eta_1}{1-\lambda\delta_h}\right) + \lambda\delta_l\left(v - \frac{\eta_1}{1-\lambda\delta_h} - \frac{\eta_2\lambda(\delta_h+\delta_l)}{(1-\lambda\delta_h-\lambda\delta_l)^2}\right)$, which after using $\delta_l = \delta_t - \delta_h$ equals $\lambda\delta_t\left(v - \frac{\eta_1}{1-\lambda\delta_h} - \frac{\eta_2\lambda\delta_t}{(1-\lambda\delta_t)^2}\right) + \frac{\eta_2\lambda^2\delta_h\delta_t}{(1-\lambda\delta_t)^2}$. For any given δ_t, the second derivative $\frac{\partial^2\pi_D}{\partial\delta_h^2} = -\frac{2\eta_1\lambda^3\delta_t}{(1-\lambda\delta_h)^3} < 0$ and the profit is strictly concave in δ_h. The first derivative $\frac{\partial\pi_D}{\partial\delta_h} = -\frac{\eta_1\lambda^2\delta_t}{(1-\lambda\delta_h)^2} + \frac{\eta_2\lambda^2\delta_t}{(1-\lambda\delta_t)^2}$ and equating it to zero gives $\delta_h = \frac{\sqrt{\eta_2}-(1-\lambda\delta_t)\sqrt{\eta_1}}{\lambda\sqrt{\eta_2}} \leq \delta_t\ \forall 0 \leq \delta_t \leq 1$. However, we need $\delta_h \geq \max(0, \delta_t - (1-q))$ and $\delta_h \leq q$ for it to be feasible, and hence $\delta_h^*(\delta_t) = \min\left(\max\left(0, \delta_t + q - 1, \frac{\sqrt{\eta_2}-(1-\lambda\delta_t)\sqrt{\eta_1}}{\lambda\sqrt{\eta_2}}\right), q\right)$. Substituting $\delta_h = \frac{\sqrt{\eta_2}-(1-\lambda\delta_t)\sqrt{\eta_1}}{\lambda\sqrt{\eta_2}}$, after some algebra, the profit $\pi_D = \lambda\delta_t\left(v - \frac{\sqrt{\eta_2}(2\sqrt{\eta_1}-\sqrt{\eta_2})}{1-\lambda\delta_t}\right)$ which is strictly concave in δ_t because $\frac{\partial^2\pi_D}{\partial\delta_t^2} = -\frac{2\lambda^2\sqrt{\eta_2}(2\sqrt{\eta_1}-\sqrt{\eta_2})}{(1-\lambda\delta_t)^3} < 0$. Similarly, when $\delta_h = 0, \delta_t + q - 1, q$, the profits are given by $\lambda\delta_t\left(v - \eta_1 - \frac{\eta_2\lambda\delta_t}{(1-\lambda\delta_t)^2}\right)$, $\lambda\delta_t\left(v - \frac{\eta_1}{1-\lambda q+\lambda-\lambda\delta_t} - \frac{\eta_2\lambda(1-q)}{(1-\lambda\delta_t)^2}\right)$, and $\lambda\delta_t\left(v - \frac{\eta_1}{1-\lambda q} - \frac{\eta_2\lambda(\delta_t-q)}{(1-\lambda\delta_t)^2}\right)$ respectively, which are all strictly concave (the second derivatives are all negative). Hence $\pi_D(\delta_h^*(\delta_t), \delta_t)$ is piecewise concave in δ_t. It is also continuous in δ_t because (i) $\pi_D(\delta_h, \delta_t)$ is continuous in (δ_h, δ_t) and (ii) $\delta_h^*(\delta_t)$ is also continuous in δ_t. So there is a unique δ_t^* that maximizes π_D. ∎

Proof (Proof of Theorem 1) The profits π_S^* and π_D^* are strictly increasing in v because an IR constraint always binds under optimality in both problems $\mathcal{S}$ and $\mathcal{D}$. From Proposition 3, we find that π_S^* is either $\left(\sqrt{v} - \sqrt{\eta_2}\right)^2$, $\lambda(1-q)\left(v - \frac{\eta_2}{1-\lambda+\lambda q}\right)$, $\left(\sqrt{v} - \sqrt{\eta_1}\right)^2$, or $\lambda\left(v - \frac{\eta_1}{1-\lambda}\right)$. All these functions are convex in v and π_S^* is continuous and differentiable in v so that π_S^* is also convex in v. For problem $\mathcal{D}$, from the proof in Proposition 4, we find that the best profit for any given δ_t is one of the following four functions (corresponding to different values of $\delta_h^*(\delta_t)$): $\lambda\delta_t\left(v - \frac{\sqrt{\eta_2}(2\sqrt{\eta_1}-\sqrt{\eta_2})}{1-\lambda\delta_t}\right)$, $\lambda\delta_t\left(v - \eta_1 - \frac{\eta_2\lambda\delta_t}{(1-\lambda\delta_t)^2}\right)$, $\lambda\delta_t\left(v - \frac{\eta_1}{1-\lambda q+\lambda-\lambda\delta_t} - \frac{\eta_2\lambda(1-q)}{(1-\lambda\delta_t)^2}\right)$, and $\lambda\delta_t\left(v - \frac{\eta_1}{1-\lambda q} - \frac{\eta_2\lambda(\delta_t-q)}{(1-\lambda\delta_t)^2}\right)$. Clearly, if $\delta_t^* = 0$ then the profit is zero (independent of v), or if $\delta_t^* = 1/\lambda(\eta_1-\eta_2)/\eta_1$ then it is linearly increasing in v; hence, it is convex in both cases. Otherwise, it is an interior solution and makes one of the derivatives of the above four functions wrt δ_t zero. Applying Envelope Theorem we find that $\frac{d\pi_D^*}{dv} = \lambda\delta_t^*(v)$, in which $\delta_t^*(v)$ is the optimal δ_t when customers' valuation is v, because the partial derivative of all the four functions wrt v is $\lambda\delta_t$. Also, $\delta_t^*(v)$ is increasing in v because (i) their derivatives wrt δ_t are increasing in v and (ii) the functions are concave in δ_t. Hence $\frac{d\pi_D^*}{dv}$ is increasing in v and so pi_D^* is convex in v. Because the two functions are convex and strictly increasing, they intersect at most at two points, which yields the result on the threshold(s) in Proposition 1. As $v \to \infty$, $\pi_S^* \to \lambda v - \frac{\eta_1\lambda}{1-\lambda}$. If $\frac{1}{\lambda}\left(1 - \frac{\eta_2}{\eta_1}\right) < 1$ then $\delta_t^* < 1\ \forall v > 0$ and so $\lim_{v\to\infty}\pi_D^* < \lambda\delta_t^* v < \lambda v - \frac{\eta_1\lambda}{1-\lambda}$. Otherwise, $\lim_{v\to\infty}\delta_t^* = 1$ and after some algebra, we obtain $\lim_{v\to\infty}\pi_D^* = \lambda v - \frac{\eta_1\lambda}{1-\lambda q} - \frac{\eta_2\lambda^2(1-q)}{(1-\lambda)^2} > \lim_{v\to\infty}\pi_S^* = \lambda v - \frac{\eta_1\lambda}{1-\lambda}$ because $\frac{\eta_2}{\eta_1} < \frac{1-\lambda}{1-\lambda q}$. ∎

Proof (Proof of Proposition 5) In the first case, $p_1 + \frac{\eta_1}{1-\lambda} \leq p_2 + \eta_1$ implies $p_1 + \frac{\eta_2}{1-\lambda} \leq p_2 + \eta_2$. Because the prices are non-negative, SP 1 benefits from satisfying more customers and so $\delta_{11}^* = q$ and $\delta_{12}^* = q$. Alternatively, in the last case the price p_1 is too high and so SP 1 cannot satisfy any customers. In the second case, we have $p_1 + \eta_1 < p_2 + \eta_1/(1-\lambda)$ (since $p_1 < p_2$) and $p_1 + \frac{\eta_1}{1-\lambda} > p_2 + \eta_1$ so that $p_1 + \eta_1/(1-\lambda\alpha_1) = p_2 + \eta_1/(1-\lambda+\lambda\alpha_1)$ and $0.5 < \alpha_1 < 1$. Also, $\alpha_2 > \alpha_1$ because either $\alpha_2 = 1$ or

$p_1+\eta_2/(1-\lambda\alpha_2) = p_2+\eta_2/(1-\lambda+\lambda\alpha_2)$. Since $p_1 < p_2$ we have $p_1+\eta_1/(1-\lambda\delta_1) \leq p_2+\eta_1/(1-\lambda+\lambda\delta_1) \Rightarrow$ $p_1 + \eta_2/(1 - \lambda\delta_1) \leq p_2 + \eta_2/(1 - \lambda + \lambda\delta_1)$ and patient customers *get a preference* over impatient ones. Under optimality SP 1 satisfies the maximum amount of customers while satisfying (12) and (13); hence $\delta_{12}^* = \min(\alpha_2, 1-q)$ and $\delta_{11}^* = (\alpha_1 - \delta_{12}^*)^+$. Similarly, in the fourth case, it can be shown that impatient customers get a preference, and the optimal fractions can be derived. Finally, in the third case, the prices are equal and so the total fraction of customers satisfied by each SP has to be 0.5. However, note that unlike the other cases, the fractions of patient and impatient customers are not unique but that does not affect the profits.

If the SPs just set the prices, we prove that self-selection by customers still results in the same fractions. It is evident in the first, third, and fifth cases in Proposition 5 because different δ_{11} or δ_{12} values result in higher dis-utility for some customers. In the second case, we first note that the total fraction of customers satisfied δ_1 ($\delta_1 = \delta_{11} + \delta_{12}$) is at least α_1 (otherwise some customers would benefit in shifting from SP 2 to SP 1). There are two possibilities: (i) $\delta_1 = \alpha_1$ which is possible only when all the patient customers who get a preference are satisfied, i.e., $\delta_{12} = 1 - q$. Further, $1 - q \leq \alpha_2$ because otherwise some patient customers would be better by shifting from SP 1 to SP 2, or (ii) $\delta_1 > \alpha_1$ in which none of the impatient customers buy from SP 1 (if any of them did, they would be better by buying from SP 2 instead) so that $\delta_{11} = 0$; the impatient customers buy from SP 1 until either all of them do so or they obtain the same dis-utility as buying from SP 2 and hence $\delta_{12} = \min(\alpha_2, 1 - q)$. Similarly, we can show that self-selection also results in the same fractions in the fourth case. ∎

Proof (Proof of Theorem 2) We use superscript $\tilde{}$ to denote the corresponding equilibrium values. First, we show by contradiction that an equilibrium, if it exists, has to be symmetric with $\tilde{p}_1 = \tilde{p}_2$ (and hence $\tilde{\alpha}_1 = \tilde{\alpha}_2 = 0.5$). Suppose $\tilde{p}_1 < \tilde{p}_2$. If $\tilde{p}_1 + \frac{\eta_1}{1-\lambda} \leq \tilde{p}_2 + \eta_1$ then $\tilde{\pi}_2 = 0$. It is not an equilibrium because SP 2 can obtain a positive profit by reducing the price from $\tilde{p}_2$ to $\tilde{p}_1$. Hence $\tilde{p}_1 + \frac{\eta_1}{1-\lambda} > \tilde{p}_2 + \eta_1$ and $\tilde{\alpha}_1 < \tilde{\alpha}_2$. Then, from Proposition 5, we know that $\tilde{\delta}_{11} = \left(\tilde{\alpha}_1 - \tilde{\delta}_{12}\right)^+$ and $\tilde{\delta}_{12} = \min(\tilde{\alpha}_2, 1-q)$. Further, either $\tilde{\alpha}_2 \leq 1-q$ or $\tilde{\alpha}_1 \geq 1-q$ because otherwise $\tilde{\alpha}_1 < 1-q < \tilde{\alpha}_2$ with $\tilde{\delta}_{11} = 0$ and $\tilde{\delta}_{12} = 1-q$ and it does not yield an equilibrium because price of SP 1 can be increased until $\alpha_1 = 1-q$ without any loss in her sales, thereby increasing her profit.

Suppose $\tilde{\alpha}_2 \leq 1-q$ then $\tilde{\delta}_{11} = 0$ and $\tilde{\delta}_{12} = \tilde{\alpha}_2$ so that $\tilde{\delta}_{21} = q$ and $\tilde{\delta}_{22} = 1 - q - \tilde{\alpha}_2$ because all the customers, due to their high valuation, buy from one of the SPs. Also, due to a result for SP 2 analogous to that in Proposition 5 (case 4), we have $\tilde{\delta}_{21} = \min(\alpha_1, q)$ and $\tilde{\delta}_{22} = \left(\tilde{\alpha}_2 - \tilde{\delta}_{21}\right)^+$. Hence $\tilde{\alpha}_2 - q = 1 - q - \tilde{\alpha}_2$ so that $\tilde{\alpha}_2 = 0.5$ which implies that $\tilde{p}_1 = \tilde{p}_2$ and $\tilde{\alpha}_1 = \tilde{\alpha}_2 = 0.5$, a contradiction. Similarly, if $\tilde{\alpha}_1 \geq 1-q$ then $\tilde{\delta}_{11} = \tilde{\alpha}_1 + q - 1$ and $\tilde{\delta}_{12} = 1-q$ so that $\tilde{\delta}_{21} = 1 - \tilde{\alpha}_1$ and $\tilde{\delta}_{22} = 0$. Also, we have $\tilde{\delta}_{21} = \min(\tilde{\alpha}_1, q)$ and $\tilde{\delta}_{22} = \left(\tilde{\alpha}_2 - \tilde{\delta}_{21}\right)^+$. Hence $\min(\tilde{\alpha}_1, q) \geq \tilde{\alpha}_2$ which implies that $\tilde{\alpha}_1 \geq \tilde{\alpha}_2$, a contradiction.

Similarly, it can be shown that $\tilde{p}_1 > \tilde{p}_2$ results in contradictions. Therefore, $\tilde{p}_1 = \tilde{p}_2$ with $\tilde{\alpha}_1 = \tilde{\alpha}_2 = \tilde{\delta}_1 = \tilde{\delta}_2 = 0.5$. If $q \geq 0.5$ and $0 < \delta_1^* < 1$ then, after some algebra, we find from Proposition 5 that (i) $p_1 < p_2$ implies $\delta_{11}^* = \alpha_1 + 1 - q$ and $\delta_{12}^* = 1 - q$ and (ii) $p_1 > p_2$ implies $\delta_{11}^* = \alpha_1$ and $\delta_{12}^* = 0$. Hence $\delta_1^* = \alpha_1$ $\forall p_1, p_2$ and $p_1 + \frac{\eta_1}{1-\lambda\delta_1^*} = p_2 + \frac{\eta_1}{1-\lambda+\lambda\delta_1^*}$. The equilibrium price is obtained by equating the derivative of $\lambda\delta_1\left(p_2 + \eta_1\left(\frac{1}{1-\lambda+\lambda\delta_1} - \frac{1}{1-\lambda\delta_1}\right)\right)$ with respect to δ_1 at $\delta_1 = 0.5$ to zero. If $q < 0.5$ then Proposition 5 implies that $\forall p_1, p_2$ s.t. $q \leq \alpha_2 \leq 1-q$ either $\delta_{11}^* = 0$ and $\delta_{12}^* = \alpha_2$ (if $p_1 < p_2$) or $\delta_{11}^* = q$ and $\delta_{12}^* = \alpha_2 - q$, and so $\delta_1^* = \alpha_2$ with $p_1 + \frac{\eta_1}{1-\lambda\delta_1^*} = p_2 + \frac{\eta_1}{1-\lambda+\lambda\delta_1^*}$. Because the equilibrium has to be symmetric, it's necessary that the derivative of $\lambda\delta_1\left(p_2 + \eta_1\left(\frac{1}{1-\lambda+\lambda\delta_1} - \frac{1}{1-\lambda\delta_1}\right)\right)$ with respect to δ_1 at $\delta_1 = 0.5$ be zero. However, that is not sufficient. An SP should also obtain a lower profit by charging a higher price and selling only to impatient customers (it is relatively straightforward to show that charging a lower price than the equilibrium price of $\frac{\eta_2\lambda}{\left(1-\frac{\lambda}{2}\right)^2}$ always reduces the profit). ∎

Proof (Proof of Proposition 6) The first and second cases, in which $\beta_1 \leq \beta_2$ and $\beta_1 > \beta_2$, are similar to to the second and fourth cases in Proposition 5 respectively. The proofs for (i) deriving the optimal values of δ_{11} and δ_{12} and (ii) showing that they result uniquely from customers' self-selection are similar to the corresponding proofs in Proposition 5. The four possibilities directly follow from the mathematical expressions for δ_{11}^* and δ_{12}^*. ∎

Proof (Proof of Theorem 3) The first part of Theorem 3 follows from Proposition 6. The case with $\tilde{\delta}_{11} > 0$ and $\tilde{\delta}_{12} = 1 - q$ does not occur at equilibrium because SP 2 would then be better by just offering single service. Next, we derive the range(s) of q for which the strategy of SP 1 is dominated under different cases. **I** Suppose SP 1 only satisfies impatient customers at equilibrium. Then her profit is $\lambda\tilde{\delta}_{11}\tilde{p}_1$. Also, due to self-selection and all the customers buying from one of the SPs, we have $\tilde{p}_1 + \frac{\eta_1}{1-\lambda\tilde{\delta}_{11}} = \tilde{p}_{2h} + \frac{\eta_1}{1-\lambda q+\lambda\tilde{\delta}_{11}}$. From optimizing the profit and equating the derivative to zero, we have $\tilde{p}_1 = \eta_1\lambda\tilde{\delta}_{11}\left(\frac{1}{(1-\lambda q+\lambda\tilde{\delta}_{11})^2} + \frac{1}{(1-\lambda\tilde{\delta}_{11})^2}\right)$. Also, note that $\tilde{p}_1 + \frac{\eta_2}{1-\lambda\tilde{\delta}_{11}} \geq \tilde{p}_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}\right)}$. We first consider the case when this inequality binds. We show that, under some conditions, the profit of SP 1 can be increased by offering differentiated service. Let $\epsilon > 0$, $p_{1h} \equiv \tilde{p}_1$, and $p_{1l} \equiv \tilde{p}_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon\right)} - \frac{\eta_2}{\left(1-\lambda\tilde{\delta}_{11}\right)\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)}$. If SP 1 charges p_{1h} and p_{1l} for high-priority and low-priority services (and the prices of SP 2 remain the same), then fractions δ_{11} and ϵ select them respectively. This strategy is feasible and profitable if and only if $\frac{\eta_2\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)^2} \leq p_{1h} - p_{1l} \leq \frac{\eta_1\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{1-\lambda\tilde{\delta}_{11}-\lambda\epsilon}$ (strong IC conditions) and $p_{1l} > 0$. We show that these conditions are satisfied for a small enough $\epsilon > 0$ if $q \leq 1/3$ or $0.5 \leq q < \hat{q}$ with $\hat{q} = \min\left(\max\{q : \frac{\lambda q}{(1-\lambda q)^3} \leq \frac{1}{(1-\lambda+\lambda q)^2}\}, 1\right)$. We find that

$$
\begin{aligned}
p_{1l} &= \tilde{p}_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon\right)} - \frac{\eta_2}{\left(1-\lambda\tilde{\delta}_{11}\right)\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)} \\
&= \tilde{p}_1 + \frac{\eta_2}{1-\lambda\tilde{\delta}_{11}} - \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}\right)} + \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon\right)} \\
&\quad - \frac{\eta_2}{\left(1-\lambda\tilde{\delta}_{11}\right)\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)} \\
&= \tilde{p}_1 + \frac{\eta_2}{1-\lambda\tilde{\delta}_{11}} - \frac{\eta_2}{(1-\lambda\tilde{\delta}_{11})(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon)} - \frac{\eta_2\lambda\epsilon}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon)} \\
&= p_{1h} - \frac{\eta_2\lambda(\tilde{\delta}_{11}+\epsilon)}{(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon)^2} + \frac{\eta_2\lambda^2\epsilon(\tilde{\delta}_{11}+\epsilon)}{(1-\lambda\tilde{\delta}_{11})(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon)^2} \\
&\quad - \frac{\eta_2\lambda\epsilon}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon)}.
\end{aligned}
$$

Hence $p_{1h} - p_{1l} \geq \frac{\eta_2\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)^2}$ for small enough $\epsilon > 0$ if $\lim_{\epsilon\to 0} \frac{\eta_2\lambda^2\epsilon(\tilde{\delta}_{11}+\epsilon)}{(1-\lambda\tilde{\delta}_{11})(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon)^2} - \frac{\eta_2\lambda\epsilon}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11}+\lambda\epsilon)} < 0$, i.e., $\frac{\lambda\tilde{\delta}_{11}}{(1-\lambda\tilde{\delta}_{11})^3} < \frac{1}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})^2}$. Because $\tilde{\delta}_{11} \leq q$ and the LHS is increasing in $\tilde{\delta}_{11}$ while the RHS is decreasing in $\tilde{\delta}_{11}$, this inequality holds if $q < \hat{q}$. Also, because $p_{1h} - p_{1l} \to \frac{\eta_2\lambda\tilde{\delta}_{11}}{\left(1-\lambda\tilde{\delta}_{11}\right)^2} < \frac{\eta_1\lambda\tilde{\delta}_{11}}{1-\lambda\tilde{\delta}_{11}}$ as $\epsilon \to 0$, we have $p_{1h} - p_{1l} \leq \frac{\eta_1\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{1-\lambda\tilde{\delta}_{11}-\lambda\epsilon}$ for small enough $\epsilon > 0$. Further, p_{1l} is positive because as $\epsilon \to 0$, $p_{1l} \to \tilde{p}_1 - \frac{\eta_1\lambda\tilde{\delta}_{11}}{1-\lambda\tilde{\delta}_{11}} > 0$.

Suppose SP 1 satisfies only impatient customers and $\tilde{p}_1 + \frac{\eta_2}{1-\lambda\tilde{\delta}_{11}} > \tilde{p}_{2l} + \frac{\eta_2}{\left(1-\lambda q+\lambda\tilde{\delta}_{11}\right)\left(1-\lambda+\lambda\tilde{\delta}_{11}\right)}$. Then $\tilde{p}_{2h} - \tilde{p}_{2l} = \frac{\eta_2\lambda(1-\delta_{11})}{(1-\lambda+\lambda\delta_{11})^2}$ because otherwise p_{2l} can be increased to increase the profit of SP 2. Also, note that if $q < \overline{q}$ then after some algebra, as shown above, $p_{1h} - p_{1l} \geq \frac{\eta_2\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)^2}$ for small enough $\epsilon > 0$. However, it is now non-trivial and we need some more properties to show that $p_{1h} - p_{1l} \leq \frac{\eta_1\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)}$ for such $\epsilon > 0$. Further, note that if this inequality holds then $p_{1l} > 0$. First, we find that $\tilde{\delta}_{11} > q/2$. That is because for any given p_1, $\pi_{2,SD} = \lambda p^*_{2h}\delta^*_{2h} + \lambda(1-q)p^*_{2l} = \lambda\left(p_1 + \frac{\eta_1}{1-\lambda q+\lambda\delta^*_{2h}} - \frac{\eta_1}{1-\lambda\delta^*_{2h}}\right)\delta^*_{2h} + \lambda(1-q)p^*_{2l}$ and since $\frac{dp^*_{2l}}{d\delta^*_{2h}} < 0$ (the optimal price p^*_{2l} is either limited by p_1 through constraint (17) or by constraint (17); both these constraints become tighter as δ_{2h} increases) and $\frac{d\pi_{2,SD}}{d\delta^*_{2h}}|_{\delta^*_{2h}=\tilde{\delta}_2 h} = 0$, we have $\tilde{p}_1 - \frac{\eta_1}{(1-\lambda\tilde{\delta}_{2h})^2} + \frac{\eta_1}{(1-\lambda\tilde{\delta}_{2h})^2} > 0$. Substituting for $\tilde{p}_1$ (see above) and using $\tilde{\delta}_{2h} = q - \tilde{\delta}_{11}$, after some algebra, we get $(1-\lambda q+\lambda\tilde{\delta}_{11})^3 > (1-\lambda q)^3$

and so $\tilde{\delta}_{11} > q/2$. The price difference for SP 1, as $\epsilon \to 0$, is then given by

$$\begin{aligned}
p_{1h} - p_{1l} &= (p_{1h} - p_{2h}) + (p_{2h} - p_{2l}) + (p_{2l} - p_{1l}) \\
&= \left(\frac{\eta_1}{1-\lambda q + \lambda\tilde{\delta}_{11}} - \frac{\eta_1}{1-\lambda\tilde{\delta}_{11}}\right) + \left(\frac{\eta_2\lambda(1-\tilde{\delta}_{11})}{(1-\lambda+\lambda\tilde{\delta}_{11})^2}\right) \\
&+ \left(\frac{\eta_2}{(1-\lambda\tilde{\delta}_{11})^2} - \frac{\eta_2}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})}\right) \\
&< \frac{\eta_2}{(1-\lambda\tilde{\delta}_{11})^2} + \frac{\eta_2\lambda(1-\tilde{\delta}_{11})}{(1-\lambda+\lambda\tilde{\delta}_{11})^2} - \frac{\eta_2}{(1-\lambda q+\lambda\tilde{\delta}_{11})(1-\lambda+\lambda\tilde{\delta}_{11})} \\
&< \frac{\eta_2}{(1-\lambda q)^2} + \frac{\eta_2\lambda(1-q/2)}{(1-\lambda+\lambda q/2)^2} - \frac{\eta_2}{1-\lambda+\lambda q},
\end{aligned}$$

in which the last inequality is due to $q/2 < \tilde{\delta}_{11} < q$. Hence if $\frac{\eta_2}{(1-\lambda q)^2} + \frac{\eta_2\lambda(1-q/2)}{(1-\lambda+\lambda q/2)^2} - \frac{\eta_2}{1-\lambda+\lambda q} \leq \frac{\eta_1\lambda q/2}{1-\lambda q/2} < \frac{\eta_1\lambda\tilde{\delta}_{11}}{1-\lambda\tilde{\delta}_{11}}$ then $\exists\epsilon > 0$ s.t. $p_{1h} - p_{1l} \leq \frac{\eta_1\lambda\left(\tilde{\delta}_{11}+\epsilon\right)}{\left(1-\lambda\tilde{\delta}_{11}-\lambda\epsilon\right)}$.

II Suppose SP 1 satisfies only patient customers. Then her profit is $\lambda\tilde{\delta}_{12}\tilde{p}_1$ and we have $\tilde{p}_1 + \frac{\eta_2}{1-\lambda\tilde{\delta}_{12}} = \tilde{p}_{2l} + \frac{\eta_2}{(1-\lambda q)(1-\lambda+\lambda\tilde{\delta}_{12})}$. Also, we have $\tilde{p}_1 + \frac{\eta_1}{1-\lambda\tilde{\delta}_{12}} \geq \tilde{p}_{2h} + \frac{\eta_1}{1-\lambda q}$. We first consider the case when this inequality binds. Let $\epsilon > 0$, new prices p_{1h} and p_{1l} be such that $p_{1h} + \frac{\eta_1}{1-\lambda\epsilon} = p_{2h} + \frac{\eta_1}{1-\lambda q+\lambda\epsilon}$ and $p_{1l} + \frac{\eta_2}{(1-\lambda\epsilon)(1-\lambda\tilde{\delta}_{12})} = p_{2l} + \frac{\eta_2}{(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})}$. With these prices, the new profit is $\lambda\epsilon p_{1h} + \lambda(\tilde{\delta}_{12} - \epsilon)p_{1l}$. Also, we have

$$\begin{aligned}
p_{1h} - p_{1l} &= (p_{1h} - p_{2h}) + (p_{2h} - \tilde{p}_1) + (\tilde{p}_1 - p_{2l}) + (p_{2l} - p_{1l}) \\
&= \frac{\eta_1}{1-\lambda q+\lambda\epsilon} - \frac{\eta_1}{1-\lambda\epsilon} + \frac{\eta_1}{1-\lambda\tilde{\delta}_{12}} - \frac{\eta_1}{1-\lambda q} + \frac{\eta_2}{(1-\lambda q)(1-\lambda+\lambda\tilde{\delta}_{12})} - \frac{\eta_2}{1-\lambda\tilde{\delta}_{12}} \\
&+ \frac{\eta_2}{(1-\lambda\epsilon)(1-\lambda\tilde{\delta}_{12})} - \frac{\eta_2}{(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})} \\
&= \frac{\eta_1\lambda\tilde{\delta}_{12}}{(1-\lambda\tilde{\delta}_{12})(1-\lambda\epsilon)} + \frac{(\eta_2-\eta_1)\lambda\epsilon}{(1-\lambda\tilde{\delta}_{12})(1-\lambda\epsilon)} + \frac{\lambda\epsilon}{(1-\lambda q)(1-\lambda q+\lambda\epsilon)}\left(\frac{\eta_2}{1-\lambda+\lambda\tilde{\delta}_{12}} - \eta_1\right) \\
&< \frac{\eta_1\lambda\tilde{\delta}_{12}}{(1-\lambda\tilde{\delta}_{12})(1-\lambda\epsilon)},
\end{aligned}$$

in which the last inequality follows from $\eta_2/\eta_1 < 1-\lambda$. Hence, $p_{1h} - p_{1l} < (\to)\frac{\eta_1\lambda\tilde{\delta}_{12}}{1-\lambda\tilde{\delta}_{12}}$ $\forall\epsilon > 0$ (as $\epsilon \to 0$) so that the IC constraints are satisfied for small enough $\epsilon > 0$. In addition to satisfying the IC constraints, the new profit has to be higher. Next, we show that if ϵ is low then $\lambda\epsilon p_{1h} + \lambda(\tilde{\delta}_{12} - \epsilon)p_{1l} \geq \lambda\tilde{\delta}_{12}\tilde{p}_1$. This inequality holds if and only if

$$\begin{aligned}
& p_{1h} - p_{1l} \geq \frac{\tilde{\delta}_{12}}{\epsilon}(\tilde{p}_1 - p_{1l}) \\
\Leftrightarrow\ & p_{1h} - p_{2h} + p_{2h} - p_{2l} + p_{2l} - p_{1l} \geq \frac{\tilde{\delta}_{12}}{\epsilon}(\tilde{p}_1 - p_{2l} + p_{2l} - p_{1l}) \\
\Leftrightarrow\ & p_{2h} - p_{2l} + \frac{\eta_1}{1-\lambda q+\lambda\epsilon} - \frac{\eta_1}{1-\lambda\epsilon} + \frac{\eta_2}{(1-\lambda\epsilon)(1-\lambda\tilde{\delta}_{12})} - \frac{\eta_2}{(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})} \\
\geq\ & \frac{\eta_2\lambda\tilde{\delta}_{12}}{(1-\lambda q)(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})} + \frac{\eta_2\lambda\tilde{\delta}_{12}}{(1-\lambda\epsilon)(1-\lambda\tilde{\delta}_{12})} \\
\Leftrightarrow\ & p_{2h} - p_{2l} + \frac{\eta_1}{1-\lambda q+\lambda\epsilon} - \frac{\eta_1-\eta_2}{1-\lambda\epsilon} - \frac{\eta_2(1-\lambda q+\lambda\tilde{\delta}_{12})}{(1-\lambda q)(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})} \geq 0 \\
\Leftarrow\ & \frac{\eta_2\lambda(1-\tilde{\delta}_{12})}{(1-\lambda+\lambda\tilde{\delta}_{12})^2} + \frac{\eta_1}{1-\lambda q+\lambda\epsilon} - \frac{\eta_1-\eta_2}{1-\lambda\epsilon} - \frac{\eta_2(1-\lambda q+\lambda\tilde{\delta}_{12})}{(1-\lambda q)(1-\lambda q+\lambda\epsilon)(1-\lambda+\lambda\tilde{\delta}_{12})} \geq 0 \\
\Leftarrow\ & \frac{\eta_2(1-\lambda q+\lambda\tilde{\delta}_{12})}{(1-\lambda q)^2(1-\lambda+\lambda\tilde{\delta}_{12})} - \frac{\eta_2\lambda(1-\tilde{\delta}_{12})}{(1-\lambda+\lambda\tilde{\delta}_{12})^2} - \eta_2 < \frac{\eta_1\lambda q}{1-\lambda q}
\end{aligned}$$

for low $\epsilon > 0$. Further, for any given q, $\frac{\eta_2(1-\lambda q+\lambda\tilde{\delta}_{12})}{(1-\lambda q)^2(1-\lambda+\lambda\tilde{\delta}_{12})}$ and $\frac{\eta_2\lambda(1-\tilde{\delta}_{12})}{(1-\lambda+\lambda\tilde{\delta}_{12})^2}$ are both strictly decreasing in $\tilde{\delta}_{12}$), and $(1-q)/2 < \tilde{\delta}_{12} < 1-q$ (the reasoning for $(1-q)/2 < \tilde{\delta}_{12}$ is similar to that of $q/2 < \tilde{\delta}_{11}$). Hence the last inequality above is satisfied $\forall\tilde{\delta}_{12}$ if $\frac{\eta_2(2+\lambda-3\lambda q)}{(1-\lambda q)^2(2-\lambda-\lambda q)} - \frac{\eta_2\lambda q}{(1-\lambda q)^2} - \eta_2 < \frac{\eta_1\lambda q}{1-\lambda q}$.

Suppose SP 1 satisfies only patient customers and $\tilde{p}_1 + \frac{\eta_1}{1-\lambda\tilde{\delta}_{12}} > \tilde{p}_{2h} + \frac{\eta_1}{1-\lambda q}$. Then $\tilde{p}_{2h} - \tilde{p}_{2l} = \frac{\eta_1\lambda(1-\tilde{\delta}_{12})}{1-\lambda+\lambda\tilde{\delta}_{12}}$; otherwise p_{2h} can be increased. It can still be shown that $p_{1h} - p_{1l} \leq \frac{\eta_1\lambda\tilde{\delta}_{12}}{1-\lambda\tilde{\delta}_{12}}$ and that the new profit is higher for some $\epsilon > 0$ if q satisfies the above inequality. However, the condition $p_{1h} - p_{1l} \geq \frac{\eta_2\lambda\tilde{\delta}_{12}}{(1-\lambda\tilde{\delta}_{12})^2}$ is no longer trivially satisfied. After some algebra, as $\epsilon \to 0$, we have $p_{1h} - p_{1l} = p_{1h} - p_{2h} + p_{2h} - p_{2l} + p_{2l} - p_{1l} = \frac{\eta_1\lambda q}{1-\lambda q} + \frac{\eta_1\lambda(1-\tilde{\delta}_{12})}{1-\lambda+\lambda\tilde{\delta}_{12}} + \frac{\eta_2}{1-\lambda\tilde{\delta}_{12}} - \frac{\eta_2}{(1-\lambda q)(1-\lambda+\lambda\tilde{\delta}_{12})}$. Because $(1-q)/2 < \tilde{\delta}_{12} < 1-q$, $p_{1h} - p_{1l} \geq \frac{\eta_2\lambda\tilde{\delta}_{12}}{(1-\lambda\tilde{\delta}_{12})^2}$ for some $\epsilon > 0$ if $\frac{\eta_1\lambda q}{1-\lambda q} + \frac{\eta_1\lambda(1-(1-q))}{1-\lambda+\lambda(1-q)} + \frac{\eta_2}{1-\lambda(1-q)/2} - \frac{\eta_2}{(1-\lambda q)(1-\lambda+\lambda(1-q)/2)} > \frac{\eta_2\lambda(1-q)}{(1-\lambda+\lambda q)^2}$, i.e., $\frac{2\eta_1\lambda q}{1-\lambda q} + \frac{2\eta_2}{2-\lambda+\lambda q} - \frac{2\eta_2}{(1-\lambda q)(2-\lambda-\lambda q)} - \frac{\eta_2\lambda(1-q)}{(1-\lambda+\lambda q)^2} > 0$. ∎

Proof (Proof of Proposition 7) After some algebra, we find that

$$\begin{aligned}
\frac{\partial\mathcal{G}}{\partial x} &= \lambda(p_{2h} - p_{2l}) + \frac{\lambda}{(1-\lambda q+\lambda x)^2}\left(\eta_1(1-\lambda q) - \frac{\eta_2(1-\lambda q+\lambda y)}{1-\lambda+\lambda y}\right) - \frac{\lambda(\eta_1-\eta_2)}{(1-\lambda x)^2},\\
\frac{\partial^2\mathcal{G}}{\partial x^2} &= \frac{-2\lambda^2}{(1-\lambda q+\lambda x)^3}\left(\eta_1(1-\lambda q) - \frac{\eta_2(1-\lambda q+\lambda y)}{1-\lambda+\lambda y}\right) - \frac{2\lambda^2(\eta_1-\eta_2)}{(1-\lambda x)^3}.
\end{aligned}$$

Also, we find that $\frac{\partial^2\mathcal{G}}{\partial x^2} < 0\ \forall 0 \leq x, y \leq 1$ because $\frac{\eta_2(1-\lambda q+\lambda y)}{1-\lambda+\lambda y}$ is decreasing in y and hence $\eta_1(1-\lambda q) - \frac{\eta_2(1-\lambda q+\lambda y)}{1-\lambda+\lambda y} \geq \eta_1(1-\lambda q) - \frac{\eta_2(1-\lambda q)}{1-\lambda} \geq 0$ as $\eta_2/\eta_1 \leq 1-\lambda$. Equating $\frac{\partial\mathcal{G}}{\partial x}$ at x^* to zero and solving for y^* gives (24). Similarly, the partial derivatives of $\mathcal{G}$ wrt y are given by

$$\begin{aligned}
\frac{\partial\mathcal{G}}{\partial y} &= \lambda p_{2l} + \frac{\eta_2(1-\lambda+\lambda x)}{1-\lambda q+\lambda x}\cdot\frac{\lambda}{(1-\lambda+\lambda y)^2} - \frac{\lambda\eta_2}{(1-\lambda y)^2},\\
\frac{\partial^2\mathcal{G}}{\partial y^2} &= -\frac{2\eta_2(1-\lambda+\lambda x)}{1-\lambda q+\lambda x}\cdot\frac{\lambda^2}{(1-\lambda+\lambda y)^3} - \frac{2\lambda^2\eta_2}{(1-\lambda y)^3},
\end{aligned}$$

and hence $\frac{\partial^2\mathcal{G}}{\partial y^2} < 0\ \forall 0 \leq x, y \leq 1$. Equating $\frac{\partial\mathcal{G}}{\partial y}$ at y^* to zero and solving for x^* gives (25). We prove that there is at most a single x^* (and hence a single $y^* = \mathcal{H}_1(x^*)$) by first showing that $\mathcal{H}_2(\mathcal{H}_1(x))$ is strictly convex and increasing over those ranges of (x, y) which can satisfy (24) and (25) with $0 \leq x, y \leq 1$. First, note that $\frac{p_{2h}-p_{2l}}{\eta_2} \leq \frac{\eta_1-\eta_2}{\eta_2(1-\lambda x)^2}$ because otherwise $\mathcal{H}_1(x) < \frac{1}{\lambda}\left(\lambda - 1 + \frac{\lambda(1-q)}{\frac{\eta_1}{\eta_2}(1-\lambda q)-1}\right) \leq \frac{1}{\lambda}\left(\lambda - 1 + \frac{\lambda(1-q)}{\frac{1-\lambda q}{1-\lambda}-1}\right) = 0$. Similarly, $\frac{p_{2l}}{\eta_2} \leq \frac{1}{(1-\lambda y)^2}$ so that $\mathcal{H}_2(y) \geq 0$[14]. $\mathcal{H}_1(x)$ is increasing because both $(1-\lambda q+\lambda x)^2$ and $\frac{\eta_1-\eta_2}{\eta_2(1-\lambda x)^2} - \frac{p_{2h}-p_{2l}}{\eta_2}$ are both non-negative and increasing in x so that the denominator in (24) is decreasing in x. In order to show strict convexity, let $\tilde{\mathcal{H}}_1(x) \equiv (1-\lambda q+\lambda x)^2\left(\frac{\eta_1-\eta_2}{\eta_2(1-\lambda x)^2} - \frac{p_{2h}-p_{2l}}{\eta_2}\right)$. Then $\tilde{\mathcal{H}}_1$ is strictly convex because it's the product of two strictly convex, increasing, and non-negative functions. Further, because $\mathcal{H}_1(x) = \frac{1}{\lambda}\left(\lambda - 1 + \frac{\lambda(1-q)}{\frac{\eta_1}{\eta_2}(1-\lambda q)-1-\tilde{\mathcal{H}}_1(x)}\right)$ is strictly convex and increasing (SCI) in $\tilde{\mathcal{H}}_1$, $\mathcal{H}_1(x)$ is SCI in x. Similarly, it can be shown that $\mathcal{H}_2(y)$ is also SCI, and hence $\mathcal{H}_2(\mathcal{H}_1(x))$ is SCI. Therefore if $\mathcal{H}_2(\mathcal{H}_1(0) < 0$ then $\exists$ at most a single x^* such that $\mathcal{H}_2(\mathcal{H}_1(x^*) = x^*$. Finally, let $y_0 \equiv \mathcal{H}_1(0) \leq \frac{1}{\lambda}\left(\lambda - 1 + \frac{1-q}{q\left(\frac{\eta_1}{\eta_2}\cdot(1-\lambda q)-(2-\lambda q)\right)}\right)$. After some algebra, we find that $\mathcal{H}_2(\mathcal{H}_1(0)) < 0$ if $\frac{(1-\lambda+\lambda y_0)^2}{(1-\lambda y_0)^2} \leq \frac{1-\lambda}{1-\lambda q}$, i.e., $\frac{1-\lambda+\lambda y_0}{1-\lambda y_0} \leq \frac{\sqrt{1-\lambda}}{\sqrt{1-\lambda q}} \Leftarrow \frac{1-q}{q\left(\frac{\eta_1}{\eta_2}\cdot(1-\lambda q)-(2-\lambda q)\right)} \leq \frac{(2-\lambda)\sqrt{1-\lambda}}{\sqrt{1-\lambda q}+\sqrt{1-\lambda}}$, which yields (26). ∎

Proof (Proof of Theorem 4) If $\mathcal{H}_2(\mathcal{H}_1(0))|_{p_{2h}=\tilde{p}_{2h}, p_{2l}=\tilde{p}_{2l}} < 0$, then prices $p_{2h} = \tilde{p}_{2h}$ and $p_{2l} = \tilde{p}_{2l}$ yield unique values of x and y that satisfy (24) and (25). Further, $x = \tilde{\delta}_{1h} = q/2$ and $y = \tilde{\delta}_{1h} + \tilde{\delta}_{1l} = 1/2$

[14] A stronger inequality needs to be satisfied for $\mathcal{H}_2$ to be non-negative but this inequality is sufficient for the proof.

do satisfy them. Because the dis-utilities from high-priority and low-priority services are equal for SPs 1 and 2, the prices $\tilde{p}_{1h} = \tilde{p}_{2h}$ and $\tilde{p}_{1l} = \tilde{p}_{2l}$ are the *best response* prices of SP 1. Similarly, $\tilde{p}_{2h}$ and $\tilde{p}_{2l}$ are SP 2's best response prices when SP 1 charges $\tilde{p}_{1h}$ and $\tilde{p}_{1l}$. Hence these prices result in an equilibrium. Further, they result in unique $\tilde{\delta}$'s. The reasoning is as follows. Because the prices at the two SPs are equal, $\tilde{\delta}_{1h} = \tilde{\delta}_{2h}$; otherwise $\tilde{\delta}_{1h} > \tilde{\delta}_{2h}$ (or $\tilde{\delta}_{1h} < \tilde{\delta}_{2h}$) and some customers would benefit from purchasing at SP 2 instead of SP 1 (SP 1 instead of SP 2). Similarly $\tilde{\delta}_{1l} = \tilde{\delta}_{2l}$ and hence $\tilde{\delta}_1 = \tilde{\delta}_2 = 0.5$. Finally, if $\frac{2\eta_2\lambda}{(2-\lambda)^2} \leq \tilde{p}_{1h} - \tilde{p}_{1l} \leq \frac{\eta_1\lambda}{(2-\lambda)}$ then the IC conditions are satisfied (note that these conditions depend only on the *total* amount of customers buying at SP 1 or SP 2, not on the individual amounts buying high-priority and low-priority services), and all the impatient (patient) customers purchase high-priority (low-priority) service thereby yielding $\tilde{\delta}_{1h} = \tilde{\delta}_{2h} = q/2$ and $\tilde{\delta}_{1l} = \tilde{\delta}_{2l} = (1-q)/2$. ∎