Winners and Losers in Priority Services

Alex Gershkov and Eyal Winter*

January 12, 2022

Abstract

We analyze the implications of introducing priority service on consumers' welfare. In monopoly markets, introducing priority service decreases the consumers' surplus unless priority service substantially enhances the set of consumers by adding ones who would not purchase the basic service otherwise. This negative effect exists despite increasing the assignment efficiency: the monopolist extracts from consumers a total payment higher than the total efficiency gain generated by the service and hence leaves consumers worse off than when no priority is offered at all. In duopoly markets with homogeneous customers the equilibrium price and customers' welfare coincide with the monopoly outcome where this monopolist faces half of the market. With heterogeneous customers as well priority reduces the aggregated consumers' welfare. Our conclusion is that in many markets in which priority service is offered priority service erects barriers to competition that are embedded in the nature of the provided service.

Keywords: Externalities, priority pricing.
JEL classification: D62, D4.

1 Introduction

By priority service (PS) we refer to the option offered by service providers to customers to purchase the right to obtain priority over regular customers. We are primarily concerned with priority queues that distort "first-come first-served" queues by serving priority customers before regular ones.

*We wish to thank Peter Cramton, Paul Klemperer, Preston McAfee, Benny Moldovanu, Jean Tirole, Alexis Walckiers and Robert Wilson for helpful comments. Gershkov wishes to thank the Israel Science Foundation for financial support. Gershkov: Department of Economics and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, and School of Economics, University of Surrey, alexg@huji.ac.il; Winter: Department of Economics and the Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, and Department of Economics, Lancaster University, eyal.winter@mail.huji.ac.il.

PS is prevalent in many industries that involve queues, but it presents a large range of consequences for those whose waiting time is reduced as well as for those whose waiting time increases. While priority boarding and priority check-in in airlines merely grants some extra convenience for consumers who purchase it, toll (fast track) roads and priority delivery of goods can often determine the value of the ride or the purchase. If we arrive at the meeting shortly before it ends, or if the suit is delivered after the wedding takes place, benefits go practically down to null. Private service providers in the health industry will often take priority patients not next in the queue into the operating room if they pay extra. Such a priority queue can easily have major health consequences for the patients who purchase it and for those who do not.[1]

Another example of priority market is front running – fees charged by financial intermediaries for faster data transmissions and execution of trade orders. It is harmful to traders who do not pay the fee and hence excluded from this priority service. While many aspects of front running are illegal, in reality preventing it is very difficult.[2]

Our objective in this paper is to study the welfare effect of priority service on consumers. Our main insight is that priority service reduces the overall consumers' welfare unless it attracts a large set of consumers to purchase the primary service relative to the case where no priority service is offered. In several important queuing environments this expansion will not take place. If the basic service is offered for free (as in national health service) or if the pricing of priority service is done separately and independently of the pricing of the basic service, or if the value of the basic service is sufficiently high relative to the cost of waiting, priority service will not expand the set of consumers who purchase the basic service. In all these environments the optimal equilibrium prices set by the service provider leave consumers worse off with priority service than without it, in spite of the fact that it generates more efficient service schedules due to consumers' heterogeneous waiting costs (patience). In fact, under certain conditions (derived later in the paper) introducing priority service can leave all consumers worse off. The type of excessive surplus extraction generated by priority service is very different from other types of surplus extractions including ones generated by price discrimination (e.g. Mussa and Rosen [33]). Firstly, it builds on the negative externalities among consumers, and the fact that the "good" called priority is less valuable the more people purchase it. Secondly, because of the negative externalities among consumers, the degree of surplus extraction is typically greater than the total efficiency gain (i.e., the reduction in overall cost of waiting under priority service relative

[1] The use of priority service (PS) in health is sometimes practiced by public health service providers (e.g. the NHS in the UK and public hospitals in Israel). Here the validity of our result about a loss of surplus for customers depends on which group of customers one refers to. If the additional revenue generated by priority service is used to improve the health service of other NHS patients, then the practice of PS gains legitimacy as it might improve the total surplus counting all NHS customers. In this case priority service can be viewed as a tax imposed on some patients to subsidize others, which might seem less controversial but not perfectly so.

[2] See Budish, Cramton and Shim [8] for discussion of economic implications of front running and possible remedy.

to a market without priority). Finally, as we shall see the excessive power of service providers remains also when we depart from the monopolistic market structure, and introduce competition. This again won't be the case with price discrimination of any degree. Priority service can be thought of as a second-degree price discrimination where consumers valuations for the two quality goods are endogenously determined in equilibrium.

To study these welfare effects we start with a simple model of priority service. Priority customers are served before regular customers. Within each of these two groups customers are served in random order. Each customer has a constant marginal waiting cost for each of the other customers standing before him/her in the queue. These costs are differential across customers and determined by a probability distribution. The service provider who knows the distribution sets the price of priority service so as to maximize its revenue.

To demonstrate the simplest manifestation of the loss of welfare due to priority service, consider two customers who purchase a certain service. Waiting to be served second costs 1 to customer 1, and 2 to customer 2. In the absence of priority service customers are served in random order. The same applies if both customers purchase priority. If only one of them purchases it, he/she is served first. Under a random order the total expected cost of waiting is $\frac{1}{2}1 + \frac{1}{2}2 = \frac{3}{2}$. Alternatively, if priority is offered to the more impatient customer for free, then the overall (expected) cost of waiting declines from 3/2 to 1. Hence, the efficiency gain of the priority queue is 1/2. However, the service provider can extract much more than 1/2 in equilibrium. Any price of less than 1 will be accepted by player 2 regardless of the other player's decision. (If player 1 purchases the priority service then player 2's willingness to pay for it is $2 - \frac{1}{2}2 = 1$, and if player 1 is not a priority customer then player 2's willingness to pay is $\frac{1}{2}2 - 0 = 1$ as well.) Similarly, any price of less than 1/2 will be accepted by player 1 regardless of the other player's decision.

If priority service does not exist customers' overall waiting disutility is 3/2, but when the service provider determines the price of priority their total disutility is[3] $1 + 1 = 2 > 3/2$. Hence not only does the service provider levy the entire efficiency gain, it also manages to extract an additional revenue of one half, making the customers jointly worse off.

Consider now a symmetric case where both customers' cost of waiting is 1; then any priority price below 1/2 yields that purchasing priority is the dominant strategy for both players. Players' waiting time in this unique equilibrium will be exactly the same as in the case where none of them purchase priority. Hence, in the unique equilibrium outcome under optimal pricing the two customers transfer a total of 1 unit of money to the service provider without getting any relief for their waiting time. If we had 100

[3] The provider either sets the price of priority to 1, and only customer 2 buys the priority service, or the provider sets the price to 1/2 and both customers buy the priority service. While the provider is indifferent between the two possibilities in this example, we assume that he sets the price to 1 and only customer 2 buys the priority service. In the second possibility the consumers' welfare loss is even greater.

customers all with a fixed marginal cost of waiting of 1, the unique equilibrium will have each of them transfer about 50 units of money to the service provider, without improving their expected waiting time relative to the situation where there is no priority service at all.

Why do priority services treat customers so badly? The answer to this question is quite simple in the case of homogeneous customers. Here, the priority service generates no value whatsoever. Relative to the case of no priority service, what it actually does is merely transferring welfare from one customer to another customer, at a price that goes wholly to the service provider, without offering any compensation to the customer who have been made worse off. With heterogeneity, the priority service generates efficiency gains. Some customers whose costs of waiting are excessively high may well be better off compared with the case of no priority service, in spite of the high price they might pay for it, but under a mild condition on the probability distribution (over the cost of waiting) the total welfare increase enjoyed by high-cost customers is offset by the price they have to pay and by the loss borne by the low-cost customers who get later service. Observe that introducing priority service diminishes the attractiveness of the regular service since customers of this service lose precedence to priority customers. Reducing the value of the regular service allows the provider to extract from the priority customers more than the increase in the efficiency. Hence priority service yields a negative total welfare for customers and a bounty for the service provider.

Our argument here as well as in the examples discussed earlier relies on the assumption that the price of the basic service remains fixed and does not decline due to the introduction of priority service. In the sequel our analysis will also address the case where the price of the basic service is endogenously determined, yet we believe that the one-dimensional pricing assumption is very reasonable in many important applications. First, we would note that any application in which priority service is offered at a price in addition to a free service offered by the government warrants analysis under such an assumption (e.g., national health systems in the UK and Israel allow priority patients to jump the queue regardless of urgency). Second, in several important applications the priority service is completely separate from the market of the primary service, either because of a price control over the latter or because it is set by a different economic entity. Some utility companies that supply a price-controlled good offer a separate service that offers priority in answering phone calls and in earlier warnings regarding power cuts (see in https://www.enwl.co.uk/power-cuts/priority-services-register/why-become-a-priority-services-customer/). Finally, SAT prep courses offered by private companies also provide a sort of priority service totally separate from the pricing of the basic service, i.e., the test itself.

As we shall see later, if the price of the basic service is endogenous but its value for consumers substantially exceeds the cost of waiting (e.g., the value of a flight vs. the cost of waiting to board the plane), the effect of priority service on consumers' welfare will still be negative. However, if the payoff of the basic service is in the same order of magnitude as the cost of waiting to the extent that priority service attracts substantially more consumers who would not purchase the basic service without it, then priority service would have a positive effect on consumers' welfare.

Our results apply in two major contexts. The first concerns a priority service offered by a service provider at a higher price than the basic service/good. The second involves environments in which a single market of public service splits as a result of introducing a secondary (private) market that offers the same service with an improved quality at a higher price (introducing or expanding private education in a way that consumes the professional resources of the free public education system is a primary example). In both contexts consumers' welfare is affected by two countervailing forces when a priority service/private service is introduced. First, efficiency gains are introduced by endogenously matching the consumers to one of the two markets based on their willingness to pay (instead of matching them randomly to a service). Second, the average quality that consumers get at the standard quality market service declines, and consumers at the improved quality market pay a higher price. The first force increases the overall welfare of consumers while the second one reduces it. Our results compare the overall welfare of consumers in these two regimes (i.e., the unified market and the split market). Our main results in both contexts show that if supply and demand are held fixed when moving from one regime to another then the split market regime generates a lower total welfare for consumers. In other words, the second force described above overrides the first one. Indeed, if supply and demand are not held fixed when moving from one regime to another, then the result can be reversed, e.g., if by abolishing priority service some impatient consumers decide to forgo consumption altogether, or if the private market of education brings about an expansion of the set of quality teachers, then the split market might introduce further benefits for consumers that exceed the downside of splitting and so splitting may be preferable.

We later study the case of multiple priority levels.[4] A customer purchasing priority service of level k is guaranteed to be served before any customer who purchased a lower priority level and after any customer who purchased a higher priority level. Customers of the same priority level are served in random order. Equilibrium selection implies that more priority levels lead to more efficient scheduling. Hence, the total welfare (of all customers and the service provider) increases with the number of levels. One should therefore hope that with a large number of priority levels customers will get in total some share of these growing efficiency gains, and be made better off compared with when there is no priority at all. This is, however, not the case. We show that consumers' welfare loss due to priority service applies to any number of levels of priority service. Interestingly, this result applies not just to the monopolist's optimal price but to every price that attracts some consumers. Furthermore, the monopolist's profit strictly increases with the number of levels. The level of surplus extraction here can be quite staggering. If, for example, the distribution of waiting costs is uniform, then the service provider's revenue can be twice as high as the total efficiency gains (relative to

[4] Using multiple priority classes is a common practice in shipping, e.g., Amazon offers standard vs. Prime two-day delivery vs. Prime one-day vs. Prime now (one- or two-hour delivery), and in visa application, e.g., the UK offers standard vs. priority vs. super-priority service.

the no-priority service case).

While most of our results are derived for the case in which the price of the primary good is fixed, in Section 7 we also study the case in which it is endogenously determined. We find that in this framework priority service can be welfare-enhancing. Such enhancement is achieved by consumers whose costs of waiting are high to the extent that without priority service they avoid buying the good altogether. Priority service significantly reduces the waiting costs for such service customers and allows them to consume the service. Our analysis in this more general framework provides sufficient conditions for the loss of welfare under the regime of priority service and even identifies the conditions under which the loss of welfare not only arises in aggregated terms but also in individual terms, i.e., conditions that imply that the priority service regime reduces the welfare of all the consumers.

We further analyze nonlinear waiting costs. Convex homogeneous waiting costs create strategic complementarities between the customers: the more customers join the priority service, the more substantial is the cost-saving from joining the priority service for other customers, and hence it is more beneficial to join the priority service. The opposite happens with concave waiting costs. For heterogeneous costs, the main result of customers' surplus extraction holds for both convex and concave waiting costs.

Priority services are mostly offered by de facto monopolies, mainly because they are almost always secondary to some other primary service (e.g., priority boarding is secondary to the flight, priority delivery is secondary to the product delivered, etc.).[5] Hence, priority services are susceptible to the holdup problem. Once a customer commits to a primary service provider he/she cannot purchase priority elsewhere. Nevertheless, our analysis here covers also the duopoly case, and reveals inherent barriers to competition in these markets. Our model of a priority service duopoly game is a simple two-stage Bertrand game. In stage 1 service providers decide simultaneously on the price of priority. In stage 2 customers' sort themselves between the two providers and between the two queues within each provider (priority and regular). The subgame-perfect equilibrium requires that no customer can be made better off by switching a provider or a queue within a provider for any prices set by the providers. Moreover, no provider can increase its revenue by changing its priority price taking into account equilibrium behavior by customers following such a change.

We show that priority service in a duopoly presents an intrinsic barrier to competition. Under homogeneity (identical costs of waiting) the duopoly does not increase competition at all relative to the case of monopoly. The unique equilibrium under optimal pricing splits the set of customers equally between the two service providers and each of the providers extracts from its set of customers exactly the same revenue that it would have extracted had it served this set of customers as a monopolist. Under heterogeneous costs we show for the parameterized class of distribution functions that even in case of duopoly competition introducing priority services decreases the aggregated consumers' welfare, in a sharp contrast to the outcome of a Bertrand competition

[5]For the effect of the hold-up problem on the pricing of ancillary goods see Gomes and Tirole [23].

for a standard good.

The intuition behind the barrier to competition that is inherent in the structure of priority service markets is quite simple. In the case of a standard product, when one provider reduces the price of the product below the price charged by its competitor, it is able to attract the other provider's customers, without fearing of losing any of its current customers. This is not necessarily the case in markets for priority services. As customers move from the more expensive provider to the less expensive one, the latter becomes more congested. As the set of priority customers grows, the priority service becomes less valuable. Some customers might prefer now to join regular service and by doing so will reduce the revenue of the competing provider. Hence, markets for priority service introduce tacit collusion that requires no communication, no signals, and not even good will – just profit maximization.

Related literature

In the classic literature on rationing and priority pricing, Wilson [36] and Chao and Wilson [11] analyze welfare-maximizing properties of priority pricing in the context of markets with random shocks like electricity provision markets. Priority pricing there is used as a rationing tool for market clearing. They show equivalence in terms of the induced allocation between welfare-maximizing priority pricing and spot pricing. In particular, they show the existence of priority pricing scheme that implements the allocation that maximizes the total welfare (of consumers and producers). Furthermore, the scheme can be adjusted for redistribution of the raised revenues among the customers such that this scheme Pareto dominates random assignment. We provide a counterweight to this important literature by showing that a profit-maximizing provider yields the customers a surplus below the surplus from the random queue. Moreover, in some cases, profit-maximizing priority service may decrease the expected utility of every customer relative to a random queue. Bulow and Klemperer [9] show conditions under which regulated prices (and the appropriate rationing) decrease consumers' surplus in competitive markets. Furthermore, Hall [24] uses a dynamic mechanism design approach to show that express toll lanes can be used in a Pareto-improving manner to deal with traffic congestion. Dworzsac, Kominers, and Akbarpour [18] addresses a more general mechanism design problem of reallocation of resources in an environment where agents differ in their valuation for an object and money.

Hassin and Haviv [26] provide an excellent survey of models on queueing. In Chapter 4 they deal with different models of priority. While they illustrate some models of monopolistic service providers, they don't illustrate the welfare impact of such policies. Moreover, they don't provide analyses of competition between the providers. Haviv and Winter [27] study a queuing model with stochastic arrival and show that the optimal pricing of priority service requires discrimination between agents even when they are identical and belong to the same priority.

Deneckere and McAfee [16] analyzed the possibility to price differentiate by introducing inferior (damaged) goods. Decreasing the quality of the basic product allows charging higher price of the better product. Notice that in our setting the value of the services is specified endogenously, in equilibrium. In some sense, introducing priority service creates "equilibrium damaged" good – because of the negative externalities the values of the damaged (regular service) and non-damaged (priority service) goods are specified by the consumers' choices in equilibrium.

Mechanism design literature on queueing started with Dolan [17].[6] This paper extends the classic characterization of Vickrey, Clark, and Groves mechanisms to the queueing environment. Follow-up analyses have considered different cost structures and evaluated their implications on implementation of the first-best efficient allocation, while satisfying budget balancedness (for a recent survey of this literature see Chun, Mitra, and Mutuswami [12]). However, this literature does not analyze the effect of priority services on customers' surplus. Glazer and Hassin [22] analyzed effect of stable transfer schemes on consumers utilities.

Another related paper is Hoppe, Moldovanu and Ozdenoren [28] who consider a two-sided market with heterogeneous, privately informed agents. Agents in the two sides of the markets are complementary to one another in generating the joint surplus. They first announce their type to a planner who then, depending on the reports, forms pairs and extracts payments. Our (monopoly) priority service model can be reformulated as a two sided market where one of the sides is passive. The main objective of Hoppe et al [28] paper is to show that assortative matching doesn't provide substantial improvement relative to coarse matching (which divides each side of the market to two sections, High and Low and then randomly matches each section in one market to the corresponding section in the other one). However that paper doesn't deal with our main issue, which is the comparison between random matching of agents to service slots (no priority service) and priority service (that can be interpreted as a coarse matching in a one-sided market).

Duopoly price competition between service providers in queueing is analyzed in Luski [31] and Levhari and Luski [30]. They analyze different models in which each provider with limited capacity decides on the price of its services and faces a stream of randomly arriving customers. The main question studied in these papers is the existence of a symmetric equilibrium in which both providers charge the same price. For analysis with more than two providers (including a continuum of providers) see Reitman [34]. These studies do not address the question of the effect of priority service on the customers. Moreover, there is no natural counterpart of this question in the models studied in these papers.

It is well known that in the case of congestion (or in the case of externality in consumption in general), the revenue-maximizing non-discriminating monopolist sets welfare-maximizing price if faced with customers with homogeneous costs; see Edelson [19]. De Borger and Van Dender [15] show that a duopoly market equilibrium with

[6] In the mechanism design literature that followed Dolan [17], the environment analyzed in this paper is called sequencing problem, while by queueing setup usually called dynamic setup with stochastic arrival of new customers.

linear demand and homogeneous costs has a higher than socially optimal congestion level.[7] Moreover, Acemoglu and Ozdaglar [1] show that increasing competition in congested markets can reduce efficiency. These papers do not address the main question of our paper, which is the effect of priority on customers' surplus. In addition, the main difference between congested markets and markets for priority is that in the models of congested markets the derived demand stems from the comparison between participation in the congested market and staying out, hence customers who stay out of the market impose no externalities, whereas in the priority markets, the derived demand for priority follows from comparison between priority and regular service, and hence the customers who do not acquire priority service impose externalities on other customers in regular service. In addition, the value of regular service is specified endogenously and indirectly in the priority markets.

For the case of a single server provider our model can be formally regarded as a model of contracting with externalities (see Segal [35] and Winter [37]). While most of this literature concern environments of complete information ours is not. Hence we prefer to think of our model as one describing markets rather than a relationship between a principal and an agent. Indeed, our section on competition cannot be embedded in the framework of contracting.

This paper is organized as follows. After presenting the basic illustration and the model in the next two sections Section 4 shows the impact of priority pricing on consumers' welfare. Section 5 extends the analysis to multiple priority classes. Section 6 generalizes the model to markets where the monopolist charges optimal prices for the standard and priority services. Section 7 shows that the main conclusion of the model remains even if the customers have nonlinear waiting costs. Competition between service providers is analyzed in Section 8. Section 9 concludes. Most proofs are presented in Appendix. Appendix A contains proofs of monopoly part, while Appendix B contains proofs of duopoly.

2 Illustration

We illustrate in two very simple examples the implication of a revenue-maximizing priority provider on customers' surplus. Assume first that there is a clientele consisting of n homogeneous customers with the same waiting costs per service normalized to 1. There is a single service provider (monopolist) with capacity normalized to 1 per period. The overall waiting cost of all customers is $\frac{n(n-1)}{2}$, and hence the aggregated utility of the customers without priority pricing is $-\frac{n(n-1)}{2}$, while the average utility is $-\frac{(n-1)}{2}$. Assume now that the provider introduces a priority service. The monopolist announces a price p. Customers, after observing the price for priority decide whether to join the priority service or to consume the regular, free service. A customer who acquires

[7] De Borger and Van Dender [15] also analyze the capacity choices of providers, which we do not address.

priority service is served before the regular customers, while within each category (both priority and regular) the service order is random.[8] Assuming n^p other customers acquire priority service, the expected utility of customer i is

$$-p - \frac{n^p}{2}$$

if i acquires priority service, and

$$-n^p - \frac{n - n^p - 1}{2} = -\frac{n + n^p - 1}{2}$$

if i doesn't acquire priority service and hence is served as a regular customer. Therefore, joining the priority service improves individual utility by $\frac{n-1}{2}$ independently of the action of the other agents. Put differently, by joining the priority service, customer i overcomes on average $\frac{n-1}{2}$ other customers independently of the size of priority and regular queues.[9] The actions of the other customers specify the composition of these $\frac{n-1}{2}$ customers, but it is irrelevant for i's decision whether to join the priority service. The next proposition characterizes the equilibrium in such markets.

Proposition 1 *In the case of homogenous customers, if $p < \frac{n-1}{2}$ all customers have a dominant strategy to join the priority service, while if $p > \frac{n-1}{2}$ all customers have a dominant strategy to join the regular, non-priority service. If $p = \frac{n-1}{2}$ all customers are indifferent between joining and not joining, independently of the choices of the other customers. The unique equilibrium outcome in the game with homogeneous customers is when the provider sets the price $p = \frac{n-1}{2}$ and all customers join the priority service.*

In this game, the monopolist is able to extract $\frac{n(n-1)}{2}$ from the customers without offering them anything since their expected waiting time remains the same. Hence the mere existence of a market for priority makes agents worse off. So the aggregated utility of the customers is given by $-\frac{n(n-1)}{2} - \frac{n(n-1)}{2}$: the aggregated waiting costs $-\frac{n(n-1)}{2}$ of the priority service exactly as without priority service, and $\frac{n(n-1)}{2}$ a total transfer to the priority service provider that corresponds to extraction of the customers' surplus.[10]

We shall now illustrate a similar result in the case where the consumers differ in their waiting costs, and hence introducing priority service increases efficiency. Assume

[8] One can think about other, more sophisticated contracts in which the monopolist applies price discrimination and makes individual offers and guarantees a specific queue position in the case of acceptance and another, very unfavorable specific position in the case of rejection. Such contracts, while theoretically interesting, are not practically appealing in many situations. Furthermore, we assume that the provider cannot artificially keep the server busy while not providing service to the waiting customers.

[9] For instance, if only one other customer joins the priority service, buying priority decreases waiting costs from $\frac{n}{2}$ to $\frac{1}{2}$, while if all other customers join the priority service, buying priority decreases the waiting time from $n-1$ to $\frac{n-1}{2}$.

[10] A similar observation appears in Hassin and Haviv [26] p. 85 and credited to a private communication with Murali Agastya from 2001.

that there is a continuum of consumers of mass 1. Consumers aren't homogeneous and face different waiting costs. Consumers' waiting costs per unit of time are given by a uniform distribution on the interval $[0,1]$. There is a single provider with capacity normalized to 1 per period. If no priority service is offered, the allocation is random, and the expected waiting time is $1/2$, which is equivalent to the aggregated consumers' surplus of $-1/4$ since the expected cost of waiting is $1/2$ per unit of time. If the provider introduces a priority service at price p, then consumers with high waiting costs pay the price p and join the priority service to save time, while the consumers with low waiting costs join the regular, free service. The utility of the consumer with waiting cost c who decides to join the regular service is given by

$$-p-c\frac{1-c^*}{2},$$

where c^* is the waiting cost of the consumer who is indifferent between these two options. If the consumer with waiting cost c joins the regular queue, his utility is $-c\left(1-\frac{c^*}{2}\right)$. As the consumer with cost c^* is indifferent between these two options, the waiting cost of this consumer depends on the price of the priority service and is given by $c^*=2p$. Moreover, the surplus of consumers with waiting costs below c^* is $-c\left(1-\frac{c^*}{2}\right)$, while the surplus of consumers with waiting costs above c^* is $-p-c\frac{1-c^*}{2}=-\frac{c^*}{2}-c\frac{1-c^*}{2}$. Hence, the aggregated consumer surplus is

$$-\frac{1}{4}-\frac{c^*}{4}+\frac{(c^*)^2}{4}.$$

The latter expression is strictly lower than $-1/4$ unless $c^*=0$ or $c^*=1$. In other words, whenever the provider offers a price such that both the regular and priority services attract a non-trivial share of consumers, the aggregate consumers' surplus decreases as a result of introducing the priority service. The revenue-maximizing provider sets $c^*=1/2$ and the customers' surplus equals $-5/16$; however, every cutoff (not only the revenue-maximizing one) decreases the consumers' surplus.

3 Model

A single provider faces a continuum of customers of mass 1.[11] Assume that customers are heterogeneous with respect to per-unit waiting time. More precisely, the distribution of customers' waiting costs per unit of time is given by distribution function F on support $[0,\overline{c}]$ with $\overline{c}<\infty$ and density $f(c)>0$ for any $c\in[0,\overline{c}]$. Hence, a customer with a per unit of time waiting cost of $c\in[0,\overline{c}]$ who gets service at time t and pays p has a utility of $-p-tc$. The provider can serve at each instant a single customer. We normalize the service time of a mass of m of consumers to be exactly m units of time, and the cost of the provider to be zero.

[11] Analyzing a continuum of customers allows us to refrain from integer problems of the queues' sizes.

While some priority services can be analyzed as completely separate, independent products (like medical procedures or loans of expensive equipment), many priority services are bundled with other products and in fact play only a secondary role in the product bundle, like priority boarding or first-class quick delivery. Hence, acquiring the major component of the bundle essentially locks the customer with the priority provider. Therefore, we start our analysis with a single service provider – a monopolist. We later extend it to a competitive environment.

We analyze a simple market interaction in which the provider at the first stage offers a non-discriminatory price p for its priority service. Customers then decide whether to acquire the offered priority service at price p, which gives them priority over regular customers. The service sequencing within each category is random.

4 Equilibrium Consumer Welfare

When the monopolist sets price p for its priority service, this price separates the clientele into two categories[12]: regular customers and priority customers. The marginal customer $c^*(p)$ is indifferent between joining the priority service and the regular one, i.e.,[13]

$$-p - c^*(p)\,\frac{1 - F(c^*(p))}{2} = -c^*(p)\left(1 - \frac{F(c^*(p))}{2}\right) \Leftrightarrow c^*(p) = 2p.$$

Therefore, if the monopolist sets a price p for its priority service with $\frac{\bar{c}}{2} \geq p \geq 0$, customers with waiting costs $c \geq c^*(p) \equiv 2p$ buy priority service and customers with waiting costs $c < c^*(p) \equiv 2p$ refrain from buying it, and consume the regular service.

To establish the effect of priority service on the customers' surplus, notice that without priority service the customers' welfare, assuming the random assignment of the queue positions, is

$$-\int_0^{\bar{c}} \frac{c}{2} f(c) dc = -\frac{\mathbb{E}(c)}{2}.$$

With priority, assuming that types above $c^*(p)$ acquire priority service and types below $c^*(p)$ get served after priority customers, the customers' welfare is[14]

$$\begin{aligned} & -\int_0^{c^*(p)} c\left(1 - F(c^*(p)) + \frac{F(c^*(p))}{2}\right) f(c)\,dc + \int_{c^*(p)}^{\bar{c}} \left(-p - c\frac{1 - F(c^*(p))}{2}\right) f(c)\,dc \\ = & -\frac{1 - F(c^*(p))}{2}\mathbb{E}(c) - \int_0^{c^*(p)} \frac{c}{2} f(c)\,dc - \int_{c^*(p)}^{\bar{c}} \frac{c^*(p)}{2} f(c)\,dc. \end{aligned}$$

[12] Depending on the price, some categories may be empty.

[13] The price of the regular service is fixed and normalized to zero. See Section 6 for discussion of the provider that sets optimal prices for both regular and priority services.

[14] Recall that to induce the division into the two categories with the indifferent type of c^*, the monopolist sets a priority price of $p(c^*) = c^*/2$.

Therefore, introduction of priority is detrimental to customers if and only if

$$F\left(c^{*}\left(p\right)\right)\mathbb{E}(c)<\int_{0}^{c^{*}(p)}cf\left(c\right)dc+c^{*}\left(p\right)\left(1-F\left(c^{*}\left(p\right)\right)\right). \tag{1}$$

Introduction of the priority service has two effects on the customers' welfare that work in opposite directions. (1) Without priority service the allocation is completely random, unrelated to the real waiting costs. From the efficiency perspective we would like to have the allocation of the slots depend on the waiting costs. Introducing a positive price for priority splits the market into two segments, whereby the customers with higher waiting costs get service first. Hence introducing priority service creates gains from the improved efficiency in allocation. McAfee [32] shows that splitting the market into two submarkets can gain a very substantial part of the fully efficient assignment.[15] (2) However, the provider can expropriate (at least part of) the created surplus via the payment for priority. The next proposition shows that if the distribution of types has an increasing failure rate (satisfies the IFR property) i.e., $\frac{1-F(c)}{f(c)}$ is decreasing, then the second effect dominates and introducing priority service is necessarily detrimental to customers' surplus. In other words, the provider extracts from the customers more than the created gains from the improved efficient allocation.

Proposition 2 *Assume that F satisfies the IFR property. Then the customers' welfare if priority service is not available is higher than if priority service is offered.*

Proof. We show that for any $c'\in(0,\overline{c})$ it must be the case that

$$F\left(c'\right)\mathbb{E}(c)<\int_{0}^{c'}cf\left(c\right)dc+c'\left(1-F\left(c'\right)\right).$$

To see this, observe that for $c'=0$ both sides of the inequality are 0, while for $c'=\overline{c}$ both sides of the inequality are equal to $\mathbb{E}(c)$. The derivative of the left-hand side of the inequality with respect to c' is $f\left(c'\right)\mathbb{E}(c)$, while the derivative of the right-hand side of the inequality with respect to c' is $c'f\left(c'\right)+1-F\left(c'\right)-c'f\left(c'\right)=1-F\left(c'\right)$. The derivative of the right-hand side is greater if and only if

$$f\left(c'\right)\mathbb{E}(c)<1-F\left(c'\right)\Longleftrightarrow\mathbb{E}(c)<\frac{1-F\left(c'\right)}{f\left(c'\right)}.$$

The IFR assumption implies that there exists c'' s.t. for all $c<c''$ the derivative of the right-hand side is greater than the derivative of the left-hand side, while for all $c>c''$ the derivative of the left-hand side is greater than the derivative of the right-hand side. ■

[15] Hoppe, Moldovanu and Ozdenoren [28] extended McAfee's analyses to markets with incomplete informations.

Remark 1 *One natural solution to the excessive market power of the provider is to restrict the number of priority slots that the provider is eligible to sell. Yet, as the last proposition shows, unless the provider is completely precluded from selling priority services such a restriction is not able to eliminate the negative effect of priority pricing on the customers' surplus. In fact the proof of the last proposition shows that introducing priority service (even without necessarily using the revenue-maximizing price) always decreases the customers' surplus if the distribution satisfies the IFR property.*

Remark 2 *The reader might find the situation of priority service to be akin to second degree price discrimination (SDPD) but the two differ in multiple aspects in terms of both intuition and formal analysis. Priority service builds on negative externalities among the customers, which plays no role in SDPD. Priority customers are offered a different good than those opting to be non-priority customers, and the willingness to pay for the improved good depends on the number of people who consume it. This dependence exacerbates the loss of surplus for customers relative to SDPD.*

In addition to analyzing the effect of priority pricing on the aggregated consumers' surplus, we can analyze its effect on individual types in terms of expected utilities (before realization of randomization of the allocation within every priority group). Clearly all types below c^* are worse off after the introduction of priority service: although these types find it optimal not to buy priority service, they must suffer from delays in getting served as now they will be served in random order after all priority customers are served. Since type c^* is indifferent between joining priority service and getting regular service, this type is worse off as well, and by continuity some types above c^* are also worse off. However, there may be types with sufficiently high waiting costs that are better off after the introduction of priority service. The utility of type $c > c^*$ before introducing priority service is $-c/2$, while after introducing the service it is

$$-p - c\frac{1 - F\left(c^*\left(p\right)\right)}{2} = -\frac{c^*\left(p\right)}{2} - c\frac{1 - F\left(c^*\left(p\right)\right)}{2}.$$

This implies that if priority service attracts a substantial mass of consumers, the time-saving element of the priority service may be limited and insufficient to compensate for the disutility from paying the price of the service. The next proposition shows the formal conditions for this comparison.

Proposition 3 *Assume that F first-order stochastically dominates the uniform distribution on $[0, \bar{c}]$. Then all types of consumers are worse off after the introduction of priority service while extreme types $\{0, \bar{c}\}$ are merely indifferent.*

Proof. All types below the cutoff c^* are worse off as they get served later. Types above c^* get the service earlier, but now pay $p = \frac{c^*}{2}$. Type $c > c^*$ is worse off if and only if

$$\begin{aligned} -\frac{c^*}{2} - c\frac{1 - F\left(c^*\right)}{2} &< -\frac{c}{2} \iff \\ cF\left(c^*\right) &< c^*. \end{aligned}$$

F first-order stochastically dominating the uniform distribution on $[0, \bar{c}]$ implies that for any c^* it holds that $F(c^*) \leq \frac{c^*}{\bar{c}}$. Hence, for any $c^* \in (0, \bar{c})$ and any $c \in (c^*, \bar{c})$, we have

$$cF(c^*) \leq c\frac{c^*}{\bar{c}} < c^*.$$

■

4.1 Private Provision of Service

In this section we show that a similar phenomenon of consumers' loss of welfare applies in a different economic environment where no priority service is introduced, but rather a better service takes resources from a common pool of resources. This situation well describes the allocation of teacher and student placements in a school system where the introduction of a private school attracts the better teachers that otherwise would remain in the public schools.

To illustrate this point we consider the following model. Assume that there exists a population of consumers with parameters/types $v \sim F[0, \bar{v}]$ that characterize their needs for service. There are k providers of the service with heterogeneous qualities where the quality of provider i is denoted by t_i when $t_1 > t_2 > \cdots > t_k > 0$. If a consumer with need v_i is allocated to a provider of quality t_j the consumer's utility is $v_i t_j$. Each provider can serve share $1/k$ of all consumers.

As a benchmark we first consider the case where there is no private provision, and hence the allocation of the consumers to the providers is random. In this case the random allocation between the consumers and the providers generates the expected welfare of consumers of $\bar{t}E[v]$ where $\bar{t} = \frac{t_1+t_2+\cdots+t_k}{k}$ is the average quality of the providers.

Now assume that the provider has quality t_1 and charges p. We can think about a monopoly firm that sells the service and is engaged in a contract with the single provider of quality t_1. Other providers remain in the public service and do not charge for the service. The capacity of each provider is unchanged and it is $1/k$ of the market. The price p is set such that the market clears, that is, a share of $1/k$ of the consumers are willing to pay the price p. Other consumers get randomly allocated providers from the set of remaining ones $t_2 > \cdots > t_k > 0$. The cutoff type v^* (the type that is indifferent between joining the private provider and remaining in the public system) satisfies

$$-p + v^* t_1 = v^* \hat{t}$$

where $\hat{t} = \frac{t_2+\cdots+t_k}{k-1}$. That is,

$$-p = v^* \left(\hat{t} - t_1\right).$$

Moreover, the market clearing implies that $1 - F(v^*) = 1/k$. The next proposition shows that the introduction of a private service that uses pool of joint resources decreases the consumers' welfare.

Proposition 4 *Assume that F satisfies IFR; then the consumers' welfare if private service is not offered is higher than if private service is offered.*

Since the technical proof and the intuition of Proposition 4 are very similar to Proposition 2, we do not replicate all the remaining results we illustrate for priority service to private service setup.

The consumers' loss of welfare that emerges in Proposition 2 and Proposition 4 relies on the assumption that the additional revenue generated by the priority service is not invested in increasing capacity. In the two applications of health services and education discussed above, the validity of this assumption is a political economy question and depends on many other factors. Nevertheless, it is reasonable to assume that the adjustment of capacity has a long-run effect and that in the short run it is more limited and constrained.

5 Multiple Priority Levels

In this section we analyze the case of multiple priority level and show that also here the seller can extract more than the total benefits he provides with the priority service. Very often providers offer more than one priority level.[16] Assume now that the provider sets k priority classes with prices $0 = p_1 < p_2 < ... < p_{k-1} < p_k$, where buying priority level l means that the customer will be served after all the customers who buy higher priority classes, $\{l+1, .., k\}$.[17] There exists the basic service which is provided for free. The customers from the same priority class are served in random order. Any list of k ordered prices divides the market into k categories (some may be empty). This division is specified by the cutoff types.[18] The cutoff type c_i, $i \in \{1, ..., k-1\}$ is indifferent between being in the class i and paying price p_i, on the one hand, and being served in priority class $i+1$ and paying price p_{i+1}, on the other. That is,

$$-p_i - \left[1 - F(c_i) + \frac{F(c_i) - F(c_{i-1})}{2}\right] c_i = -p_{i+1} - \left[1 - F(c_{i+1}) + \frac{F(c_{i+1}) - F(c_i)}{2}\right] c_i,$$

which implies that[19]

$$p_{i+1} = p_i + \frac{F(c_{i+1}) - F(c_{i-1})}{2} c_i.$$

This equation illustrates the trade-off that the marginal customer with type c_i faces when contemplating what priority level to choose (either level i or $i+1$). The additional costs that this customer has to pay to be upgraded to the higher priority level is $p_{i+1} - p_i$, and the time saving associated with this upgrade is $\frac{F(c_{i+1}) - F(c_{i-1})}{2}$.

[16] In addition to regular and priority service, providers often also offer super-priority service that gives priority over all other categories.

[17] Restricting attention to strictly monotone prices is without loss, as if there exists l such that $p_l \leq p_{l-1}$, then it means that at least one category is empty and there exists a sequence of monotone prices (with less categoies) that generates the same allocation and the same utilities to all participants.

[18] Choosing cutoffs is equivalent to choosing quantities. It is well known that there is no difference between a monopolist that optimally chooses prices and a monopolist that optimally chooses quantities.

[19] We use notation of $c_0 = 0$, and $c_k = \bar{c}$.

This recursive specification of prices allows us to write them as

$$p_i = \sum_{j=1}^{i-1} \frac{F(c_{j+1}) - F(c_{j-1})}{2} c_j. \tag{2}$$

Due to this recursive structure the change in the cutoff c_i affects the prices of all higher priority categories. More precisely, assume that the cutoff of category i, c_i decreases by ϵ to $c_i - \epsilon$. It switches some customers from priority class i to a higher class $i+1$. Such a change increases the size of priority class $i+1$ and so all customers in this, now larger class are willing to pay less than with the cutoff c_i. This shift decreases the size of priority class i. Moreover, since more customers now belong to the higher priority classes than with cutoff c_i, customers in priority class i are willing to pay less than before the shift. However, the increased size of class $i+1$ allows the provider to charge class $i+2$ a higher amount as the option to join class $i+1$ has become less valuable. Such a domino effect recursively influences the prices of all the higher priority classes. In the optimal mechanism the monopolist chooses cutoffs that exactly balance these two effects, i.e., the decrease in the revenues from priority class $i+1$ and the increase in the revenues from customers of higher priority classes.

The provider's revenues are given by[20]

$$\begin{aligned} R &= \sum_{i=2}^{k} p_i \left[F(c_i) - F(c_{i-1})\right] \\ &= \sum_{i=2}^{k} \left[F(c_i) - F(c_{i-1})\right] \sum_{j=1}^{i-1} \frac{F(c_{j+1}) - F(c_{j-1})}{2} c_j = \sum_{i=1}^{k-1} (1 - F(c_i)) \frac{F(c_{i+1}) - F(c_{i-1})}{2} c_i. \end{aligned}$$

One may expect that when the number of priority classes is large, then since the resulting allocation is more efficient than in case of random allocation, consumers are able to keep part of the increase in the total efficiency. However, as our next result shows, this is not the case.

Proposition 5 *Assume that F satisfies the IFR property. Then for any number of priority classes $k > 1$, the customers' welfare if priority service is not offered is higher than if priority service is offered.*

The proof of the last proposition is recursive: for any allocation cutoffs with $l-1$ priority classes $\{c_1, ...c_{l-2}\}$ when $c_{l-2} < \bar{c}$, adding another priority class that splits the highest interval of waiting costs $[c_{l-2}, \bar{c}]$ into two, i.e., having l priority classes with allocation cutoffs $\{c_1, ...c_{l-1}\}$, always decreases the aggregated consumers' surplus. Repeating this argument k times, each time adding another priority class, implies the result.

The previous proposition does not use the profit maximization arguments of the provider. However, as the next proposition shows, the provider's profits are strictly increasing in the number of classes.

[20] In the following expression we use notation $c_0 = 0$ and $c_k = \bar{c}$.

Proposition 6 *Assume that F satisfies the IFR property; then, the provider's profits are strictly increasing in the number of priority classes.*

The next example shows the optimal mechanism in the case of k priority classes where the waiting costs are distributed according to the uniform distribution on $[0, 1]$.

Example 1 *Assume a uniform distribution of waiting costs with support* $[0, 1]$. *If there are k categories, the optimal price for priority category $i \in \{1, ..., k\}$ is*

$$p_i = \frac{i(i-1)}{2k^2}.$$

The provider's optimal revenues and the corresponding customers' surplus are

$$\begin{aligned} R &= \frac{1}{6}\left[1 - \frac{1}{k^2}\right], \\ CS &= -\frac{1}{3} + \frac{1}{12k^2}. \end{aligned}$$

Therefore, in the limit (when $k \to \infty$) the customers' surplus goes to $-1/3$. Recall that without any priority, the customers' surplus was $-\mathbb{E}(c)/2 = -1/4$ (which corresponds to $k = 1$ in the expressions above). Hence in the limit, the total efficiency gain in the equilibrium allocation is $1/12$ *(the total waiting costs decrease from $-1/4$ to $-1/6$) while the service provider gets twice! as much revenue* $(1/6)$.

6 Endogenous Pricing of the Basic Product

We have assumed thus far that the price of the basic service for which priority is offered is exogenous and fixed. In the Introduction we listed several important priority markets in which this is the case and argued that this assumption holds to a certain extent also in other markets where the booking procedure separates the decision regarding the basic service from the decision regarding priority. Nevertheless, we provide here also the analysis of the case where the price of the basic service is determined endogenously by the monopoly, and identify conditions under which our main finding applies here as well.

For simplicity, we remain within the one-dimensional framework; i.e., we assume heterogeneity of types only in terms of their waiting costs. We denote by V the consumers' value from the service. As it turns out, the distinction between the cases in which priority service decreases consumers' welfare and the cases in which it increases it is already apparent under this simplifying assumption. We normalize the utility of the customers from not consuming the good to 0. There is a continuum of customers of mass 1 who differ in terms of their waiting costs. We adopt the assumptions that were introduced in Section 3 regarding the distribution of the individual waiting costs.

The utility of a customer with a waiting cost of c per unit of time if he gets service in period t and pays p is $V - ct - p$.

We start our comparison with the case of a provider that offers regular service only and chooses its price optimally, to maximize its revenue. Denote by p its price. Assume that consumers with waiting costs below $c^r \in [0, \bar{c}]$ join the queue while the rest do not.

$$\underset{0}{|}\underbrace{- - - - - - - - - - -}_{\text{regular service}}\underset{c^r}{|}\underbrace{- - - - - - - - - - -}_{\text{no service}}\underset{1}{|}$$

The utility of consumers with waiting costs c from joining the service is given by

$$V - p - c\frac{F\left(c^r\right)}{2}.$$

Since the utility from not joining is 0, the marginal type c^r should be indifferent between joining the queue and remaining unserved; that is c^r satisfies

$$V - p - c^r \frac{F\left(c^r\right)}{2} = 0.$$

Notice that a revenue-maximizing monopolist will never set up a price below $V - \bar{c}\frac{F(\bar{c})}{2} = V - \frac{\bar{c}}{2}$ since it would attract the same share of consumers as the price $V - \frac{\bar{c}}{2}$.

The seller's revenue is $pF(c^r)$, while the consumers' surplus is

$$\int_0^{c^r} \left(V - p - c\frac{F\left(c^r\right)}{2}\right) f(c)dc.$$

Assume now that the provider sets two types of services: the regular service and the priority service. Therefore, the provider sets two prices: one for the regular service and one for the priority service. Consumers with very low waiting costs choose the regular service, consumers with very high waiting costs do not join any service, while consumers in the middle range join the priority service.

$$\underset{0}{|}\underbrace{- - - - - - - - - -}_{\text{regular service}}\underset{c^I}{|}\underbrace{- - - - - - - - - -}_{\text{priority service}}\underset{c^{II}}{|}\underbrace{- - - - - - - - -}_{\text{no service}}\underset{1}{|}$$

Denote by Π the price of the priority service and by π the price of the regular service with $\Pi \geq \pi$. The utility of consumers with waiting costs c from joining the regular service is

$$V - \pi - c\left[\frac{F\left(c^I\right)}{2} + F\left(c^{II}\right) - F\left(c^I\right)\right] = V - \pi - c\left[F\left(c^{II}\right) - \frac{F\left(c^I\right)}{2}\right]$$

while the utility of consumers with waiting costs c from joining the priority service is

$$V - \Pi - c\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}.$$

Type c^I is indifferent between getting the regular service and the priority service, while type c^{II} is indifferent between getting the priority service and no service at all. That is,

$$\begin{aligned} V - \pi - c^I \left[F\left(c^{II}\right) - \frac{F\left(c^I\right)}{2} \right] &= V - \Pi - c^I \frac{F\left(c^{II}\right) - F\left(c^I\right)}{2} \Leftrightarrow \\ \Pi - \pi &= c^I \frac{F\left(c^{II}\right)}{2} \end{aligned} \tag{3}$$

and

$$\begin{aligned} V - \Pi - c^{II} \frac{F\left(c^{II}\right) - F\left(c^I\right)}{2} = 0 \Leftrightarrow \\ \Pi = V - c^{II} \frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}. \end{aligned} \tag{4}$$

The provider's expected revenue is

$$R = \pi F\left(c^I\right) + \Pi \left[F\left(c^{II}\right) - F\left(c^I\right) \right]$$

while the consumers' surplus is

$$\int_0^{c^I} \left(V - \pi - c \left[F\left(c^{II}\right) - \frac{F\left(c^I\right)}{2} \right] \right) f(c) dc + \int_{c^I}^{c^{II}} \left(V - \Pi - c \frac{F\left(c^{II}\right) - F\left(c^I\right)}{2} \right) f(c) dc.$$

We first show that introducing priority service always (strictly) increases the provider's revenues.

Lemma 1 *Introducing priority service increases the provider's revenues.*

The next proposition extends the main result to the setting where the monopolist sets optimal prices for the regular service and the priority service. However, now we need a slightly different assumption on the distribution of types.

Definition 1 *Distribution function F satisfies a decreasing reversed failure rate (DRFR) if $F(c)/f(c)$ is increasing in*[21] $c \in [0, \bar{c}]$.

Proposition 7 *Assume that F satisfies DRFR. Moreover, assume that $(V - \bar{c}) \geq \frac{1}{2f(\bar{c})}$. Then the customers' welfare under a monopolistic regime that offers only regular service is higher than the customers' welfare if the monopolist offers both regular and priority services.*

[21] Since $\left(\frac{f(c)}{F(c)}\right)' = (\ln F(c))''$, DRFR is equivalent to the log-concavity of F. See Bagnoli and Bergstrom [6] for a discussion of log-concave distribution functions.

The conditions of the last proposition guarantee that the provider optimally sets $c^r = c^{II} = \bar{c}$; in other words, if the consumers assign a very high value to the service, relative to their waiting costs, then it is not profitable for the monopolist to exclude any consumer. To see this observe that if $c^r < \bar{c}$ then increasing c^r has two effects on the provider's revenue. On the one hand, it decreases the price $p = V - c^r \frac{F(c^r)}{2}$, but on the other hand it increases the set of consumers that get the service and pay p. Condition $(V - \bar{c}) \geq \frac{1}{2f(\bar{c})}$ together with DRFR guarantees that the second effect dominates and the provider sets the highest price that includes all the customers. The intuition for $c^{II} = \bar{c}$ is similar.

The main difference between this case and the result in Proposition 2 is that, in the present setting, introduction of the priority service allows the price of the regular service to be adjusted. In other words, if the priority service is not available, the price of the regular service is dictated by the participation constraint of the most impatient customer – the customer with waiting costs $\bar{c}$. If the priority service is offered, the price of the priority service is given by the participation of this type and the size of the priority queue, while the price of the regular service is given by the indifference condition of the cutoff type c^I. Hence, in the regime without priority service, the provider sets the price of its service to

$$p = V - \frac{\bar{c}}{2},$$

while in the case where both the regular service and the priority service are provided, the prices of the services are

$$\begin{aligned}
\Pi &= V - \bar{c}\frac{1 - F\left(c^I\right)}{2} = V - \frac{\bar{c}}{2} + \bar{c}\frac{F\left(c^I\right)}{2} \\
\pi &= \Pi - \frac{c^I}{2} = V - \frac{\bar{c}}{2} + \frac{\bar{c}F\left(c^I\right) - c^I}{2}.
\end{aligned}$$

While for any $c^I \in (0, \bar{c})$ we have $\Pi > p$, the comparison between p and π is less clear as the next example illustrates. We may have cases where $\pi > p$, $\pi < p$, and $\pi = p$.

Example 2 *1. Assume that the costs of waiting are distributed according to function $F(c) = \sqrt{c}$ for $c \in [0, 1]$, and the value of the service is $V = 2$. Then the optimal price if only the regular service is provided is $p = 3/2$. If the provider offers both the priority service and the regular service, the optimal prices are (it is optimal for the provider to set $c^I = 1/3$)*

$$\begin{aligned}
\Pi &= V - \bar{c}\frac{1 - F\left(c^I\right)}{2} = \frac{3}{2} + \frac{\sqrt{1/3}}{2} \approx 1.79 \\
\pi &= V - \bar{c}\frac{1 - F\left(c^I\right)}{2} - \frac{c^I}{2} = \frac{3}{2} + \frac{\sqrt{1/3} - 1/3}{2} \approx 1.62 > p.
\end{aligned}$$

2. Assume that the costs of waiting are distributed according to function $F(c) = c$ for $c \in [0, 1]$, and the value of the service is $V = 3/2$. Then the optimal price if only *the regular service is provided is $p = 1$. If the provider offers both the priority service and the regular service, the optimal prices are (it is optimal for the provider to set $c^I = 1/2$)*

$$\Pi = V - \bar{c}\frac{1 - F\left(c^I\right)}{2} = 1 + \frac{1/2}{2} = 1.25$$
$$\pi = V - \bar{c}\frac{1 - F\left(c^I\right)}{2} - \frac{c^I}{2} = 1 = p.$$

3. Assume that the costs of waiting are distributed according to function $F(c) = c^2$ for $c \in [0, 1]$, and the value of the service is $V = 5/4$. Then the optimal price if only the regular service is provided is $p = 3/4$. If the provider offers both the priority service and the regular service, the optimal prices are (it is optimal for the provider to set $c^I = 2/3$)

$$\Pi = V - \bar{c}\frac{1 - F\left(c^I\right)}{2} = \frac{3}{4} + \frac{(2/3)^2}{2} \approx 0.972$$
$$\pi = V - \bar{c}\frac{1 - F\left(c^I\right)}{2} - \frac{c^I}{2} = \frac{3}{4} + \frac{(2/3)^2 - 2/3}{2} \approx 0.639 < p.$$

The above example allows us to make the following claim.

Proposition 8 *Assume that F satisfies DRFR, and F is first-order stochastically dominated by the uniform distribution with support $[0, \bar{c}]$ and $(V - \bar{c}) \geq \frac{1}{2f(\bar{c})}$. Then $\Pi > \pi \geq p$ and introducing priority service makes all consumers except the type $\bar{c}$ strictly worse off.*

Proposition 7 is based on the assumption that $(V - \bar{c}) \geq \frac{1}{2f(\bar{c})}$, which guarantees that the value of the service is high enough that the provider optimally does not exclude any consumer. However, if the value of the service is not high enough, then the priority service has a consumer surplus-enhancing effect. In the next lemma we show that if $(V - \bar{c}) < \frac{1}{2f(\bar{c})}$, then introducing priority service increases the set of consumers that get service.

Lemma 2 *Assume that F satisfies DRFR. Moreover, assume that $(V - \bar{c}) < \frac{1}{2f(\bar{c})}$. Then $c^r < c^{II} \leq \bar{c}$, i.e., introducing priority service decreases the set of consumers excluded from the service.*

For instance, if $V = 1$ and the costs of waiting are distributed according to the uniform distribution on $[0, 1]$ then $c^r = \sqrt{\frac{2}{3}} \approx 0.816$, while $c^{II} = \frac{\sqrt{2}}{3} \approx 0.943$. Hence types in the interval (c^r, c^{II}) are excluded from the market in case no priority service is offered (and get zero utility), but participate in the trade and obtain positive utility if the provider offers both regular and priority services.[22] Observe that the last lemma

[22] See the proof of Proposition 9 for the derivations of the optimal cutoffs for this distribution.

does not imply that $(V - \bar{c}) < \frac{1}{2f(\bar{c})}$ leads necessarily to superiority of the regime with priority service. Rather, it implies the existence of the additional effect of extending the set of consumers that favor the regime with priority service.

We show in the next proposition that for a large class of parameterized distributions (that satisfy the DRFR property), introducing priority service increases the aggregated consumers' surplus. The novel effect that changes the comparison stems from the observation that introducing priority service reduces the set of excluded consumers. That is, introducing priority service attracts types that otherwise would remain outside of the market. For this effect we need V to be not very high, relative to $\bar{c}$.

Proposition 9 *Assume that the waiting costs are distributed according to distribution function* $F(c) = c^{\theta}$ *where* $\theta > 0$ *and* $c \in [0, 1]$, *and the value of the service is* $0 < V \leq 1$. *Then the aggregated consumers' surplus is higher in the case where the priority service is offered.*

The general insight of this section is that in economic environments where the prices of both the basic service and the priority service are endogenously determined, priority service can be welfare-enhancing for consumers if it allows the set of consumers to substantially expand by including also consumers who would not purchase the basic service without it. Such environments require the value of the primary good/service to be at the same order of magnitude as the cost of waiting. Services that give priority to "the right to choose," such as services that allow customers to choose apartments in a multi-unit buildings, are good examples for such economic environments. Ashenfelter and Genesove [5] analyze auction data in such markets, and a random priority for the right to choose has also been used recently in Israel for its affordable housing program. It is reasonable to expect that priority service in such environments enhances customers' welfare, as the value of the primary good is closely entangled with the cost of choosing late. Customers who have strong preferences for specific units may reasonably avoid purchasing an apartment in such buildings unless they are offered priority in choosing their desired unit.

Another possible application in which the cost of waiting can be a major fraction of the basic service is that of priority service in the cargo shipping industry. Shipping companies offer such a service to consumers by guaranteeing priority in the unloading of cargo. To the extent that some clients that purchase this priority service have pressing contractual obligations that reduce the value of the cargo dramatically, it would be reasonable to think of this market as one in which the cost is sufficiently high so that without priority service the shipping would not take place. In such an environment again priority service is likely to have positive effect on consumers' welfare.

7 Non-linear Cost

The utility function of the customers as presented in our model is derived from the customers' waiting costs, which are assumed to be linear in the waiting time. In this section we show that our results are robust to this assumption by studying the cases in which these waiting costs are either concave or convex. A priori it is not clear which of these two scenarios is more adequate for representing queueing disutility. On the one hand, cost functions in economics are typically assumed to be convex. However, the convexity of a cost function is very intuitive only in the context of production as it reflects the idea that low-hanging fruits that are picked first are less costly to pick than those that remain after the early ones are gone. In queues, however, the shape of the cost function depends either on individuals' mental discomfort from waiting, or on the way opportunities disappear due to delays. It seems to us that one can come up with arguments in favor of both concavity and convexity in both interpretations. Hence we consider both options.

We assume that waiting t units of time creates disutility of $g(t) \geq 0$, where $g(t)$ is increasing, bounded and differentiable. We first analyze the homogeneous customers. There is mass of size 1 of customers. We allow the disutility function g to be either convex or concave and we further assume that $g(0) = 0$. Assume that the provider sets a price of p for priority. If share s of the customers joins the priority service, the expected utility of a marginal customer in the priority service is

$$-p - \frac{\int_0^s g(t)\, dt}{s},$$

while the expected utility of the marginal customer in the regular service is

$$-\frac{\int_s^1 g(t)\, dt}{1-s}.$$

In case of linear waiting costs, the benefits from joining the priority service are independent of the number of other customers who join the service, and it is the reason for customers to have a dominant action. This is not the case with nonlinear costs. While joining the priority service still improves the averaged position by $1/2$ independently of the action of the other customers, in the case of nonlinear value of time the effect of such an improvement on a customer's utility depends on the actions of the other customers (or on the sizes of priority queue and regular queue). The benefits from joining the priority service if share s of customers joins are

$$B(s) = -\frac{\int_0^s g(t)\, dt}{s} + \frac{\int_s^1 g(t)\, dt}{1-s}.$$

Lemma 3 *Assume that g is concave, then $B(s)$ is decreasing. If g is convex, then $B(s)$ is increasing.*

Hence, for concave g the benefits from joining decrease with the share of customers that join the priority service. Therefore, for p such that[23]

$$B(0) \geq p \geq B(1),$$

[23] We define $B(0) = \lim_{s\to 0} B(s) = \int_0^1 g(t)\, dt$ and $B(1) = \lim_{s\to 1} B(s) = g(1) - \int_0^1 g(t)\, dt$.

a share s of customers that satisfies

$$p = \frac{\int_s^1 g(t)\,dt}{1-s} - \frac{\int_0^s g(t)\,dt}{s} \tag{5}$$

join the priority service. The next proposition generalizes the characterization of the equilibrium of linear case to concave and convex waiting costs.

Proposition 10 *Assume that g is concave, then in a unique subgame perfect equilibrium outcome all customers join the priority service and the optimal price is*

$$p^* = g(1) - \int_0^1 g(t)\,dt.$$

Assume that g is convex; then the optimal price is $p^* = g(1) - \int_0^1 g(t)\,dt$ and all customers join the priority service.

At the optimal price, similarly to the linear cost with homogeneous case, all customers join the priority service and hence the total waiting cost of these customers is the same as that of the customers without priority service, and so the transfer to the service provider is just the customers' surplus extraction.

There is a substantial difference between the convex and concave waiting cost cases. While in the case of concave g the equilibrium is unique and is in (weakly) dominant strategies, in the case of convex waiting costs the more customers that join the priority service, the more substantial is time-saving effect on the utility from joining it. Therefore, for convex g, if $p < B(0)$, it is optimal for all customers to join the priority service, and if $p > B(1)$ it is optimal for customers to use the regular service. For $p \in (B(0), B(1))$ we have three possible equilibria: (1) all customers join priority service, (2) all customers join regular service, (3) a share s of customers where s satisfy

$$p = \frac{\int_s^1 g(t)\,dt}{1-s} - \frac{\int_0^s g(t)\,dt}{s}$$

join the priority service. Adopting the standard assumption of mechanism design for choosing the best equilibrium for the seller/mechanism designer allows us to conclude that the provider will set the price of $g(1) - \int_0^1 g(t)\,dt$, which is the highest price that attracts all customers to the priority service.

We introduce heterogeneity of the weighting costs into the nonlinear cost model by assuming a specific functional form of the cost. Assume that the disutility from waiting t periods of time is ct^θ where c is the individual cost parameter that follows the assumptions introduced in Section 3 and $\theta > 0$. For different values of θ this functional form allows for both convexity and concavity of the cost in waiting time. Similar to the linear cost case, by setting the price of p for its priority service, the provider divides the customers into two categories: customers with a higher cost parameter who join priority service and customers with a lower cost parameter who join the regular service.

For a given price p, the expected utility of type c from joining priority service if all types with parameters $c \geq c^*$ join the priority service is given by

$$-p - c\frac{\int_0^{1-F(c^*)} t^\theta dt}{1 - F(c^*)} = -p - c\frac{(1 - F(c^*))^\theta}{\theta + 1},$$

while that type's expected utility from the regular service is

$$-c\frac{\int_{1-F(c^*)}^1 t^\theta dt}{F(c^*)} = -c\frac{1 - (1 - F(c^*))^{\theta+1}}{(\theta + 1) F(c^*)}.$$

Therefore, for a given price p, types above c^* join the priority service, while types below c^* join the regular service, where c^* solves[24]

$$\begin{aligned} p &= \frac{c^*}{\theta + 1}\left(\frac{1 - (1 - F(c^*))^{\theta+1}}{F(c^*)} - (1 - F(c^*))^\theta\right) \\ &= \frac{c^*}{\theta + 1}\frac{1 - (1 - F(c^*))^\theta}{F(c^*)}. \end{aligned}$$

The provider's problem is to choose the cutoff that maximizes its expected profits:

$$\max_{c^o \in [0, \bar{c}]} c^o \frac{1 - F(c^o)}{F(c^o)} \frac{1 - (1 - F(c^o))^\theta}{\theta + 1}.$$

The next proposition generalizes Proposition 2 to nonlinear waiting costs.

Proposition 11 *Assume that F satisfies the IFR property. Assume further that $\theta > 0$. Then the customers' welfare if priority service is not available is higher than if priority service is offered.*

Similarly to the linear case described in Proposition 2, the last proposition holds for any price such that the two classes are nonempty, and not necessarily the optimal price.

8 Competition

As we argued earlier, the market of priority is best described as a monopoly due to the hold-up problem arising from the fact that priority is typically offered by the same provider who provides the primary good to which the customer has already committed him/herself. Yet the exploitative nature of priority service presents itself also in a more competitive environment. To show this, we will now study a model of competition

[24] While it may not follow immediately from the equation below, there is at most one indifference type for any price.

between two identical providers. Each provider is able to serve the entire market and has a cost normalized to zero. We shall show that with homogeneous costs not only it is the case that priority service reduces customers' welfare but also that, in spite of the fact that providers can compete over the price of priority, the equilibrium price ends up being identical to the monopoly price (with the appropriate adjustment for the increase in market service capacity). In the case of heterogeneous costs, we show that for a large class of the distribution functions for any possible prices of the priority these providers can set, introducing priority service reduces customers' welfare. We further show in examples that even outside of this class of the distributions, the equilibrium prices of priority service reduce customers' welfare. We assume that two providers at the first stage simultaneously choose prices p_1 and p_2 for their priority services. At the second stage customers decide whether to join a queue of provider 1 or provider 2 and whether to buy the priority service of that provider or to get the regular service. Our assumption regarding the duopoly market is that the price of the primary service (for which priority is offered) is already fixed and identical for the two providers. This assumption can be interpreted as the outcome of a Bertrand competition over the primary service. It is also very relevant for services that are offered for free or at a fixed, regulated price (such as health insurance under national schemes) and customers pay only for add-ons and priority services.

8.1 Homogeneous Costs

We first assume that all customers have the same (linear) waiting costs, normalized to 1. Denote by $n_i^p(p_1, p_2)$ the share of customers who acquire priority services from service provider i if the prices of the providers for their priority services are p_1 and p_2, and denote by $n_i^{np}(p_1, p_2)$ the share of customers who join the regular service of provider i. The total share of customers of provider i is $n_i = n_i^p + n_i^{np}$, where $n_1 + n_2 = 1$. Like in the monopoly case, if

$$p_i \leq \frac{n_i}{2},$$

then the customers of provider i prefer priority service to regular service of that provider. Unlike the monopoly case, however, each customer has more options, as he may join the service of the other provider.

We show that in a unique pure strategy subgame-perfect equilibrium the providers set the prices of $\left(\frac{1}{4}, \frac{1}{4}\right)$ and the customers are divided such that $n_1^p = n_2^p = 1/2$. Thus, in equilibrium all customers get the priority service and each provider essentially gets the monopoly profits from half of the market.

Proposition 12 *In a unique pure strategy subgame-perfect equilibrium, prices are* $\left(\frac{1}{4}, \frac{1}{4}\right)$ *and the customers are divided such that* $n_1^p = n_2^p = 1/2$.

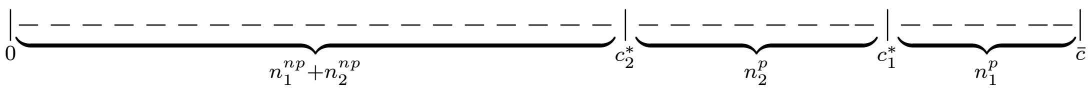
Our proof of the last proposition is insightful as it reveals the forces that cripple competition in a market of priority service. In contrast to a standard good in a market of priority service, a price cut does not guarantee a provider a larger clientele as the expansion of the clientele will make the service less attractive to existing customers who might prefer to move to the regular service and save on priority charges. More specifically, at the first stage of the proof we show how the clientele chooses between the providers and between their services for any possible pair of prices, p_1 and p_2. At the second stage we derive the optimal responses of each provider and characterize the equilibrium prices.

It is interesting to notice that although the service providers compete à la Bertrand, the equilibrium outcome is very different from the standard Bertrand competition with a perfectly competitive price. While increasing the price above 1/4 would lead to losing all priority customers similar to a Bertrand competition, decreasing the price would not attract all the customers to that provider, and this deviation would lead to an outcome very different from that of the standard Bertrand competition. Imagine that provider 1 deviates and undercuts its competitor by offering a price of $\frac{1}{4} - \varepsilon$. Observe that if in this case some share of customers move from provider 2 to the cheaper provider 1, it would take away all the priority service customers of provider 2 – as now it would hold that $p_2 > \frac{n_2}{2}$ as $p_2 = \frac{1}{4}$ and $n_2 < \frac{1}{2}$ – leaving provider 2 with only regular customers. In such a case provider 1 cannot serve in its priority service even half of the market, and, clearly, having only regular customers at provider 2 and only priority customers at provider 1 is not part of a subgame equilibrium**.** In the proof of the above proposition we show that undercutting the competitor and charging a price of $\frac{1}{4} - \varepsilon$ **does not change** the priority clientele of that provider and only causes a division of the competitor's clientele into priority and regular customers such that each customer gets exactly the same utility (as a priority customer at either of the providers or as a regular customer of provider 2). This argument essentially shows that prices $\left(\frac{1}{4}, \frac{1}{4}\right)$ are part of an equilibrium. The formal proof will show that it is the unique one.

8.2 Heterogeneous Costs

Now we assume that the two competitive providers are facing heterogeneous customers with different linear waiting costs. We adopt the assumptions that were introduced in Section 3 regarding the distribution of the individual waiting costs. Denote by p_1 and p_2 the prices for the priority service of providers 1 and 2, respectively. Assume without loss of generality that at the first stage the providers set prices such that $p_1 \geq p_2$. We first characterize the customers' optimal choices for a given pair of priority prices (p_1, p_2). We then analyze the optimal prices set by the providers.

Assume the following equilibrium structure: share $n_1^p(p_1, p_2)$ of customers with the highest waiting costs (with cost parameters between $c_1^*(p_1, p_2)$ and $\bar{c}$) choose priority service from provider 1, share $n_2^p(p_1, p_2)$ of customers with relatively high costs (with cost parameters between $c_2^*(p_1, p_2)$ and $c_1^*(p_1, p_2)$) choose priority service from provider 2 with $n_1^p \leq n_2^p$. Moreover, share $n_1^{np}(p_1, p_2)$ of customers with waiting costs below $c_2^*(p_1, p_2)$ get regular service from provider 1, and share $n_2^{np}(p_1, p_2)$ of customers with waiting costs below $c_2^*(p_1, p_2)$ get regular service from provider 2.



For any p_1 and p_2 with $p_1 \geq p_2$ the customers' decisions satisfy the following conditions:

(1) A customer with a waiting cost of c_1^* is indifferent between provider 1's and 2's priority service[25]:

$$-p_1 - c_1^* \frac{1 - F(c_1^*)}{2} = -p_2 - c_1^* \frac{F(c_1^*) - F(c_2^*)}{2}.$$

(2) A customer with a waiting cost of c_2^* is indifferent between provider 2's priority service and any regular service

$$-p_2 - c_2^* \frac{F(c_1^*) - F(c_2^*)}{2} = -c_2^* \left[F(c_1^*) - F(c_2^*) + \frac{F(c_2^*) - n_1^{np}}{2} \right].$$

(3) The expected waiting time in both providers' regular service is the same:

$$1 - F(c_1^*) + \frac{n_1^{np}}{2} = F(c_1^*) - F(c_2^*) + \frac{F(c_2^*) - n_1^{np}}{2}.$$

The last condition implies that

$$n_1^{np} = 2F(c_1^*) - 1 - \frac{F(c_2^*)}{2}.$$

Plugging it into condition (2) gives us

$$p_2 = c_2^* \left[\frac{1 - F(c_1^*)}{2} + \frac{F(c_2^*)}{4} \right].$$

By condition (1) we have

$$\begin{aligned} p_1 &= p_2 + c_1^* \left[\frac{F(c_1^*) - F(c_2^*)}{2} - \frac{1 - F(c_1^*)}{2} \right] \\ &= \frac{c_2^* - c_1^*}{2} + \left(c_1^* - \frac{c_2^*}{2} \right) \left[F(c_1^*) - \frac{F(c_2^*)}{2} \right]. \end{aligned}$$

Hence, for any $p_1 \geq p_2$ the market for priority services will be divided as follows: customers with waiting costs above c_1^* join priority service of provider 1, customers

[25] In the case of a symmetric equilibrium with $p_1 = p_2$, this condition implies that $1 - F(c_1^*) = F(c_1^*) - F(c_2^*)$. That is, each provider has the same share of priority customers.

with waiting costs in the interval $[c_2^*, c_1^*]$ join the priority service of provider 2, where c_1^* and c_2^* solve

$$p_1 - p_2 = c_1^* \left[\frac{F(c_1^*) - F(c_2^*)}{2} - \frac{1 - F(c_1^*)}{2} \right] \tag{6}$$

and

$$p_2 = c_2^* \left[\frac{1 - F(c_1^*)}{2} + \frac{F(c_2^*)}{4} \right]. \tag{7}$$

Lemma 4 *For any $p_1 \geq p_2$, there exists a unique equilibrium division of the customers into priority and regular customers; i.e., there exist unique $c_1^* \geq c_2^* \geq 0$ that satisfy requirements (1)–(3).*

The customers' selection of services that is given in (6) and (7) allows us to conclude that a zero price cannot be a part of the equilibrium strategy.

Proposition 13 *There is no pure strategy equilibrium with $p_i = 0$ for $i \in \{1, 2\}$.*[26]

The reason that equilibrium prices in both duopoly models (with both homogeneous and heterogeneous customers) are very different from the standard competitive equilibrium in the Bertrand competition is the negative externalities that customers impose on each other. By joining a queue, each customer prolongs the waiting time of all customers in this queue. These negative externalities generate market power to the service providers, which reduces competition: in the duopoly competition case (in both cases of homogeneous and heterogeneous customers), charging a lower price cannot increase the priority customer base substantially, since in the case of an increase in the number of priority customers, this service becomes less valuable to other customers.

We show next that in a big, parameterized class of distribution functions competition over priority services makes customers' welfare lower than completely random allocation - without any priority service. Observe that if $p_1 \geq p_2 > 0$ then the customers' welfare in the market with priorities is given by

$$\begin{aligned} &- \int_{c_1^*}^{\bar{c}} \left(p_1 + c \frac{1 - F(c_1^*)}{2} \right) f(c) dc - \int_{c_2^*}^{c_1^*} \left(p_2 + c \frac{F(c_1^*) - F(c_2^*)}{2} \right) f(c) dc \\ &- \int_0^{c_2^*} c \left(1 - F(c_1^*) + \frac{n_1^{np}}{2} \right) f(c) dc, \end{aligned}$$

while if no priority service is offered it is $-\frac{\mathbb{E}(c)}{4}$. That is, despite the competition, even in case of heterogeneous customers the providers are able to extract from the consumers payments in excess of the increase in the allocative efficiency.

[26] This argument applies also to the setup in which the competitors set the prices of both of their services endogenously.

Proposition 14 *Assume that $F(c) = c^\theta$ with $\theta \geq 1$. Then the customers' welfare if priority service is not available is higher than if priority service is offered by the providers.*

For $\theta \geq 1$ any positive prices for priority service decrease the consumers surplus, not only the optimal, equilibrium prices. However, as our next example shows for $\theta < 1$ the equilibrium prices are such that the consumers' welfare is lower than in the market without priority service. For its derivation see Appendix B.

Example 3 *Assume now that the distribution of types is given by $F(c) = \sqrt{c}$ for $c \in [0,1]$. Then the equilibrium cutoffs are $c_1 = 0.67336$ and $c_2 = 0.34744$. Hence, the providers set prices $p_1 = c_2\left(\frac{1-\sqrt{c_1}}{2} + \frac{\sqrt{c_2}}{4}\right) + c_1\frac{2\sqrt{c_1}-\sqrt{c_2}-1}{2} = 0.0998$ and $p_2 = c_2\left(\frac{1-\sqrt{c_1}}{2} + \frac{\sqrt{c_2}}{4}\right) = 0.08237$. The sets of priority customers of both providers are $n_1^p = 0.17941$ and $n_2^p = 0.23114$, respectively. The profits of providers 1 and 2 are $\pi_1 = 0.0179$ and $\pi_2 = 0.019$, respectively. Customers' surplus equals to -0.0878, while if priority service is not available customers' surplus is*

$$-\frac{\int_0^1 \frac{1}{2}\sqrt{s}ds}{4} = -0.083.$$

Hence introducing priority service lowers customers' surplus.

9 Discussion

Without substantially expanding the set of consumers priority service allows a service provider to extract in revenue more than the overall benefits that customers can acquire from such a regime, yielding an overall net loss to the customers. It also harms competition between service providers due to an implicit deterrence mechanism whereby attracting more customers by a price reduction may create congestion that will induce other customers to leave. Finally, it increases inequality as the benefits that priority customers purchase from the service provider for money generate no cost whatsoever to the service provider. Instead, these costs are entirely borne by another group of customers, those who fail to purchase priority.

Most of our results build on the assumption that the price of the basic service is fixed, and that the pricing game involves only the priority. Several real-life markets are consistent with this assumption. One of the most controversial ones among these, in terms of its social consequences, is the market of priority services in the health sector, where the basic good is given for free and payments apply to expediting treatment.[27]

[27] There are other similar markets including (1) Toll Roads (2) Visa priority service (UK) https://ukvisa.blog/2018/02/24/uk-priority-visa/ (3) Practitioners Priority service offered by the IRS https://www.irs.gov/tax-professionals/practitioner-priority-service-r (4) the recent Heathrow priority service that offer fast track security check for merely 12 pounds. https://www.heathrow.com/at-the-airport/airport-services/fast-track.

In the Introduction we discussed additional examples. However, in some other markets, and prominently in the airline industry, priority is sold together with a basic service (an airline ticket) by the same firm and in a more competitive environment, allowing the firm to simultaneously optimize on the two prices (of the basic service and the priority). The relevance of our model in these cases depends on the level of linkage between these two services from the customers' point of view. To be more specific, consider the process of the online booking of an airline ticket (e.g., through Expedia or Kayak). Such a process facilitates a market where the price of the basic service is practically fixed. The two factors airline customers pay most attention to when booking are the airfare and the flight's itinerary i.e., departure time and connections.[28] These and only these details are listed on the booking site before customers make their choice of flight. Only after choosing their option and going through the ticketing process they are presented with the airline priority options. Indeed, they can still opt out and check the cost of similar priority options with other airlines. However, we believe (though we have no data to support it) that by the time the choice of the ticket has already been made, and effort has been exerted toward completing the booking, it is very likely that customers' decisions are reduced to whether or not to buy priority with the airline they are already booking with, rather than continue searching.[29]

Some service providers operate multiple servers simultaneously. Under this regime our results are extendable to the case of multiple servers in a straightforward manner. In such cases it is optimal to have each server dealing with both priority and non-priority customers, handling priority customers before non-priority ones. Sometimes a service provider will dedicate one or more servers to deal with priority customers only. Such regimes are suboptimal.[30] They are used primarily because identifying the right next customer in line to be served is either impossible or costly (toll roads is a good example, where the nature of the service requires a dedicated service for priority customers). While such regimes may leave more surplus at the hand of customers they are still exposed to the same forces of exploitation that our model has.

There are several reasons why regulators might find intervention in priority service markets more justified than in many other markets that present impediments to competition. Firstly, unlike markets where the impediment to competition is caused by an exogenous constraint, such as capacity constraints (see Compte, Jenny, and Rey [13]), in PS markets it is generated by a trading practice of the firms themselves (similarly to the trading practice of exclusive dealing that is prohibited by law (see Bernheim and Whinston [7]). However, in contrast to exclusive dealing, we don't need to prohibit PS, for, as we shall argue later, milder remedies might suffice. Secondly, as we have

[28] See https://www.statista.com/statistics/428703/most-important-factors-for-choosing-flights-among-air-travelers-us/

[29] Einav et al [20] finds that eBay buyers do not fully internalize shipping fees, which are displayed after they click through to purchase.

[30] It is suboptimal also from a social welfare perspective i.e., service providers and customers, because it potentially allows a non-priority customer whose cost of waiting is low to be served before a priority one with a much higher cost of waiting.

shown, PS facilitates collusion in oligopolistic markets of priority service. Collusion outcomes tend to induce major social costs (see Harrington [25]), but in the case of PS the collusion is tacit and hence legal, unless legal constraints are imposed on how PS is allowed to be used. Thirdly, PS provides negative incentives for innovation. The general effect of barriers to competition on investment is in general quite ambiguous (see Aghion, Harris, and Vickers [4], Aghion, Harris, Howitt, and Vickers [3] and Aghion, Dewatripont, and Rey [2]), but in our framework it is more straightforward: if firms can make money out of selling PS to customers who suffer by enduring long and slow lines, and can sustain this extra revenue through a tacit collusion, why should they invest in innovation that might reduce the waiting cost of their regular service?

Designing regulatory policy to diminish the illnesses of priority service markets is challenging, and this is not the prime purpose of our paper. Some measures may turn out to be ineffective: A cap on priority service prices, for example, would hardly be successful. It will clearly increase their demand and hence impose harsher consequences on a smaller group of customers who can't even afford to pay the capped price. An alternative policy would be to limit the number of customers who receive priority service. This policy would guarantee a cap on the welfare loss of regular customers at a loss of some gains from trade in priority rights. But neither of these policies can guarantee that priority service will enhance the overall customers' surplus and, as we show, under a mild assumption on the distribution of the individual costs, they will not. The one type of policy that can guarantee it is one that facilitates trade in priority rights. In fact, such a policy guarantees not only that the total aggregate net surplus from the priority service is positive but also that priority service constitutes a Pareto improvement, i.e., by making all customers and the service provider better off relative to the benchmark of no priority service. To facilitate such a policy, it will be required that the service provider price only the regular service. All upgrades to priority service would then be auctioned out with an incentive-compatible mechanism that imposes monetary transfers between customers whereby priority customers compensate regular ones (see Kittsteiner and Moldovanu [29]). It would make sense for this service to be offered by a service provider different from the primary one, i.e., a firm that specializes in allocating priority rights and offers multiple services. The outcome of such mechanism will determine who gets priority, the price of priority and also the compensation that regular customers receive for enduring a longer line.

We would also like to stress that we are not suggesting that priority service or the privatization of a public service should be necessarily avoided even when our results point to the existence of a negative effect of priority service. The contribution of this paper for public policy is highlighting the fact that the loss of consumers' welfare due to priority service can be more prevalent than we tend to think, and that it is important to identify the losers and winners when implementing such a regime. If this is done and the revenue generated by priority service is at least partially used to compensate losers, then priority service is totally legitimate even under the conditions where we show it to be welfare-reducing for consumers. Gershkov and Schweinzer [21] showed when such "ideal" solution exists for linear waiting costs using equivalence between queueing problem and partnership resolution problem (see Cramton et al. [14]).

10 Appendix

10.1 Appendix A. Monopoly

Proof of Proposition 4. If private service is offered, then the consumers' welfare is given by

$$\begin{aligned}
& \int_{v^*}^{\bar{v}} \left(-p + v t_1\right) f\left(v\right) dv + \int_0^{v^*} v \hat{t} f(v) dv \\
= \; & v^* \left(\hat{t} - t_1\right) \left(1 - F\left(v^*\right)\right) + t_1 \int_{v^*}^{\bar{v}} v f\left(v\right) dv + \hat{t} \int_0^{v^*} v f(v) dv \\
= \; & v^* \left(\hat{t} - t_1\right) \left(1 - F\left(v^*\right)\right) + \left(\hat{t} + t_1 - \hat{t}\right) \int_{v^*}^{\bar{v}} v f\left(v\right) dv + \hat{t} \int_0^{v^*} v f(v) dv \\
= \; & v^* \left(\hat{t} - t_1\right) \left(1 - F\left(v^*\right)\right) + \left(t_1 - \hat{t}\right) \int_{v^*}^{\bar{v}} v f\left(v\right) dv + \hat{t} \mathbb{E}\left[v\right] \\
= \; & \left(t_1 - \hat{t}\right) \int_{v^*}^{\bar{v}} \left(v - v^*\right) f\left(v\right) dv + \hat{t} \mathbb{E}\left[v\right].
\end{aligned}$$

We compare it to the consumers' welfare if private service is not available which is given by $\bar{t}\mathbb{E}\left[v\right]$. The regime without private system generates higher consumers' welfare if and only if

$$\begin{aligned}
\left(t_1 - \hat{t}\right) \int_{v^*}^{\bar{v}} \left(v - v^*\right) f\left(v\right) dv + \hat{t} \mathbb{E}\left[v\right] & \leq \bar{t} \mathbb{E}\left[v\right] \Longleftrightarrow \\
\left(t_1 - \hat{t}\right) \int_{v^*}^{\bar{v}} \left(v - v^*\right) f\left(v\right) dv & \leq \left(\bar{t} - \hat{t}\right) \mathbb{E}\left[v\right]
\end{aligned}$$

Observe that

$$\begin{aligned}
\bar{t} - \hat{t} & = \frac{t_1 + t_2 + \cdots + t_k}{k} - \frac{t_2 + \cdots + t_k}{k-1} = \frac{(k-1)t_1 - t_2 - \cdots - t_k}{k\left(k-1\right)} \\
t_1 - \hat{t} & = t_1 - \frac{t_2 + \cdots + t_k}{k-1} = \frac{(k-1)t_1 - t_2 - \cdots - t_k}{k-1}.
\end{aligned}$$

Hence we can rewrite the last inequality as follows

$$\begin{aligned}
\int_{v^*}^{\bar{v}} \left(v - v^*\right) f\left(v\right) dv & \leq \frac{1}{k} \mathbb{E}\left[v\right] \Longleftrightarrow \\
-\left(1 - F(v^*)\right) \mathbb{E}\left[v\right] & \leq -\int_{v^*}^{\bar{v}} \left(v - v^*\right) f\left(v\right) dv.
\end{aligned}$$

For $v^* = 0$ and for $v^* = \bar{v}$ the last inequality holds as equality. Taking derivative of the left hand side wrt v^* we have $f(v^*)\mathbb{E}[v]$, the derivative of the right hand side is $\int_{v^*}^{\bar{v}} f(v)\,dv = 1 - F(v^*)$. The derivative of the right hand side is greater if and only if

$$f(v^*)\mathbb{E}[v] < 1 - F(v^*) \iff \mathbb{E}[v] < \frac{1 - F(v)}{f(v)}.$$

The IFR assumption implies that there exists v' s.t. for all $v < v'$ the derivative of the right hand side is greater than the derivative of the left hand side, while for all $v > v'$ the derivative of the left hand side is greater than the derivative of the right hand side.
■

Proof of Proposition 5. Our proof is recursive. We will show that for any number of priority classes l and cutoffs $\{c_1, ..., c_{l-1}\}$ when $c_{l-1} < \bar{c}$, adding another (higher) priority class and cutoff $c_l \in (c_{l-1}, \bar{c})$ decreases the consumers' welfare. Since we showed that adding a single priority class decreases the consumers' welfare, using this argument repeatedly (while each time adding another, higher priority class) allows us to conclude that any selling procedure that involves $k > 1$ priority classes with cutoffs $\{c_1, ..., c_{k-1}\}$ is inferior to no priority from the consumer welfare perspective.[31]

We will show now that adding a higher priority class decreases the consumers' welfare. If there are l priority classes with cutoffs $\{c_1, ..., c_{l-1}\}$ and $c_{l-1} < \bar{c}$, then the prices of p_l and p_{l-1} are such that the type c_{l-1} is indifferent between joining either of the two highest classes – either priority class l or $l-1$. That is,

$$\begin{aligned} -p_l - \frac{1 - F(c_{l-1})}{2} c_{l-1} &= -p_{l-1} - \left(1 - F(c_{l-1}) + \frac{F(c_{l-1}) - F(c_{l-2})}{2}\right) c_{l-1} \iff \\ p_l &= p_{l-1} + \frac{1 - F(c_{l-2})}{2} c_{l-1}. \end{aligned}$$

The consumers' welfare is given by

$$\begin{aligned} CS(&\leq c_{l-1}) + \int_{c_{l-1}}^{\bar{c}} \left[-p_l - \frac{1 - F(c_{l-1})}{2} c\right] f(c)dc \\ &= CS(\leq c_{l-1}) - p_{l-1}[1 - F(c_{l-1})] - \frac{1 - F(c_{l-2})}{2} c_{l-1}[1 - F(c_{l-1})] \\ &\quad - \frac{1 - F(c_{l-1})}{2} \int_{c_{l-1}}^{\bar{c}} cf(c)dc \end{aligned}$$

where $CS(\leq c_{l-1})$ is the consumers' welfare of all types lower than c_{l-1}.

If there are $l+1$ priority classes with cutoffs $\{c_1, ..., c_{l-1}, c_l\}$ and $c_l \in (c_{l-1}, \bar{c})$ (so that the additional class partitions the interval of the previously highest class into two intervals), then the price p'_l is such that type c_{l-1} is indifferent between priority classes

[31] If $c_{l-1} = \bar{c}$, then applying this construction to the highest cutoff below $\bar{c}$ allows us to make the same conclusion.

l and $l-1$, while the price p'_{l+1} is such that type c_l is indifferent between priority classes l and $l+1$. Observe that this change does not affect the prices of the lower classes. The first indifference condition is

$$\begin{aligned} -p_{l-1} - \left(1 - F\left(c_{l-1}\right) + \frac{F\left(c_{l-1}\right) - F\left(c_{l-2}\right)}{2}\right) c_{l-1} &= -p'_l - \left(1 - F\left(c_l\right) + \frac{F\left(c_l\right) - F\left(c_{l-1}\right)}{2}\right) c_{l-1} \Longleftrightarrow \\ p'_l &= p_{l-1} + \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1}. \end{aligned}$$

Type c_k must be indifferent between classes k and $k+1$:

$$-p'_l - \left(1 - F\left(c_l\right) + \frac{F\left(c_l\right) - F\left(c_{l-1}\right)}{2}\right) c_l = -p'_{l+1} - \frac{1 - F\left(c_l\right)}{2} c_l \Longleftrightarrow$$
$$p'_{l+1} = p'_l + \frac{1 - F\left(c_{l-1}\right)}{2} c_l = p_{l-1} + \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1} + \frac{1 - F\left(c_{l-1}\right)}{2} c_l$$

The consumers' welfare is

$$\begin{aligned} & CS(\leq c_{l-1}) + \int_{c_{l-1}}^{c_l} \left[-p'_l - \left(1 - F\left(c_l\right) + \frac{F\left(c_l\right) - F\left(c_{l-1}\right)}{2}\right) c\right] f(c)dc \\ & \quad + \int_{c_l}^{\bar{c}} \left[-p'_{l+1} - \frac{1 - F\left(c_l\right)}{2} c\right] f(c)dc \\ = {} & CS(\leq c_{l-1}) + \int_{c_{l-1}}^{c_l} \left[-p_{l-1} - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1} - \left(1 - F\left(c_l\right) + \frac{F\left(c_l\right) - F\left(c_{l-1}\right)}{2}\right) c\right] f(c)dc \\ & + \int_{c_l}^{\bar{c}} \left[-p_{l-1} - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1} - \frac{1 - F\left(c_{l-1}\right)}{2} c_l - \frac{1 - F\left(c_l\right)}{2} c\right] f(c)dc \\ & = CS(\leq c_{l-1}) - p_{l-1}\left[1 - F\left(c_{l-1}\right)\right] - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1}\left[F\left(c_l\right) - F\left(c_{l-1}\right)\right] \\ & - \left(1 - \frac{F\left(c_l\right)}{2} - \frac{F\left(c_{l-1}\right)}{2}\right) \int_{c_{l-1}}^{c_l} cf(c)dc - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1}\left[1 - F\left(c_l\right)\right] \\ & \quad - \frac{1 - F\left(c_{l-1}\right)}{2} c_l\left[1 - F\left(c_l\right)\right] - \frac{1 - F\left(c_l\right)}{2} \int_{c_l}^{\bar{c}} cf(c)dc \\ & = CS(\leq c_{l-1}) - p_{l-1}\left[1 - F\left(c_{l-1}\right)\right] - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1}\left[1 - F\left(c_{l-1}\right)\right] \\ & - \left(\frac{1 - F\left(c_l\right)}{2} + \frac{1 - F\left(c_{l-1}\right)}{2}\right) \int_{c_{l-1}}^{c_l} cf(c)dc - \frac{1 - F\left(c_{l-1}\right)}{2} c_l\left[1 - F\left(c_l\right)\right] - \frac{1 - F\left(c_l\right)}{2} \int_{c_l}^{\bar{c}} cf(c)dc \\ = {} & CS(\leq c_{l-1}) - p_{l-1}\left[1 - F\left(c_{l-1}\right)\right] - \frac{F\left(c_l\right) - F\left(c_{l-2}\right)}{2} c_{l-1}\left[1 - F\left(c_{l-1}\right)\right] - \frac{1 - F\left(c_{l-1}\right)}{2} \int_{c_{l-1}}^{\bar{c}} cf(c)dc \\ & - \frac{1 - F\left(c_l\right)}{2} \int_{c_{l-1}}^{c_l} cf(c)dc - \frac{1 - F\left(c_{l-1}\right)}{2} c_l\left[1 - F\left(c_l\right)\right] - \frac{F\left(c_{l-1}\right) - F\left(c_l\right)}{2} \int_{c_l}^{\bar{c}} cf(c)dc \end{aligned}$$

Introducing an additional priority class decreases the consumers' surplus if and only if

$$CS(\leq c_{l-1}) - p_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-1}\right)}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc >$$

$$CS(\leq c_{l-1}) - p_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{F\left(c_l\right)-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-1}\right)}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc$$

$$-\frac{1-F\left(c_l\right)}{2}\int_{c_{l-1}}^{c_l} cf(c)dc - \frac{1-F\left(c_{l-1}\right)}{2}c_l\left[1-F\left(c_l\right)\right] + \frac{F\left(c_l\right)-F\left(c_{l-1}\right)}{2}\int_{c_l}^{\overline{c}} cf(c)dc \Leftrightarrow$$

$$-\frac{1-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right] > -\frac{F\left(c_l\right)-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right]$$

$$-\frac{1-F\left(c_l\right)}{2}\int_{c_{l-1}}^{c_l} cf(c)dc - \frac{1-F\left(c_{l-1}\right)}{2}c_l\left[1-F\left(c_l\right)\right] + \frac{F\left(c_l\right)-F\left(c_{l-1}\right)}{2}\int_{c_l}^{\overline{c}} cf(c)dc.$$

Observe that for either $c_l = c_{l-1}$ or $c_l = \overline{c}$ we have equality in the previous expression. The derivative of the right-hand side of the last inequality with respect to c_l equals

$$\begin{aligned}
& -f\left(c_l\right)c_{l-1}\frac{1-F\left(c_{l-1}\right)}{2} + \frac{f\left(c_l\right)}{2}\int_{c_{l-1}}^{c_l} cf(c)dc - \frac{1-F\left(c_l\right)}{2}c_l f\left(c_l\right) - \frac{1-F\left(c_{l-1}\right)}{2}\left[1-F\left(c_l\right)\right] \\
& +\frac{1-F\left(c_{l-1}\right)}{2}c_l f\left(c_l\right) + \frac{f\left(c_l\right)}{2}\int_{c_l}^{\overline{c}} cf(c)dc - \frac{F\left(c_l\right)-F\left(c_{l-1}\right)}{2}c_l f\left(c_l\right) \\
= \; & -f\left(c_l\right)c_{l-1}\frac{1-F\left(c_{l-1}\right)}{2} + \frac{f\left(c_l\right)}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc - \frac{1-F\left(c_{l-1}\right)}{2}\left[1-F\left(c_l\right)\right] \\
= \; & f\left(c_l\right)\left[-c_{l-1}\frac{1-F\left(c_{l-1}\right)}{2} + \frac{1}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc - \frac{1-F\left(c_{l-1}\right)}{2}\frac{1-F\left(c_l\right)}{f\left(c_l\right)}\right],
\end{aligned}$$

while the derivative of the left-hand side is zero. Observe that IFR implies that the derivative that we calculated in the last expression changes its sign once from negative to positive, and hence for any $c_l \in (c_{l-1}, \overline{c})$ it holds that

$$CS(\leq c_{l-1}) - p_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-1}\right)}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc >$$

$$CS(\leq c_{l-1}) - p_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{F\left(c_l\right)-F\left(c_{l-2}\right)}{2}c_{l-1}\left[1-F\left(c_{l-1}\right)\right] - \frac{1-F\left(c_{l-1}\right)}{2}\int_{c_{l-1}}^{\overline{c}} cf(c)dc.$$

■

Derivations for Example 1. The first-order condition of maximizing the monopolist's revenue with respect to c_i for $i \in \{c_1, ..., c_k\}$ is

$$\begin{aligned}
& \frac{\left(1-F\left(c_i\right)\right)\left(F\left(c_{i+1}\right)-F\left(c_{i-1}\right)\right)}{2} - f\left(c_i\right)\frac{F\left(c_{i+1}\right)-F\left(c_{i-1}\right)}{2}c_i \\
& +f\left(c_i\right)\frac{1-F\left(c_{i-1}\right)}{2}c_{i-1} - f\left(c_i\right)\frac{1-F\left(c_{i+1}\right)}{2}c_{i+1} = 0
\end{aligned}$$

We can rewrite it as follows

$$(1 - F(c_i)) = f(c_i)\, c_i - f(c_i)\, \frac{1 - F(c_{i-1})}{F(c_{i+1}) - F(c_{i-1})} c_{i-1} + f(c_i)\, \frac{1 - F(c_{i+1})}{F(c_{i+1}) - F(c_{i-1})} c_{i+1}$$

For the uniform distribution with $[0,1]$ support we get

$$1 - c_i = c_i - \frac{1 - c_{i-1}}{c_{i+1} - c_{i-1}} c_{i-1} + \frac{1 - c_{i+1}}{c_{i+1} - c_{i-1}} c_{i+1}.$$

The solution to the last equations is

$$c_i = \frac{i}{k}$$

The optimal prices are

$$p_i = \sum_{j=1}^{i-1} \frac{F(c_{j+1}) - F(c_{j-1})}{2} c_j = \sum_{j=1}^{i-1} \frac{c_{j+1} - c_{j-1}}{2} c_j = \sum_{j=1}^{i-1} \frac{i}{k^2} = \frac{i(i-1)}{2k^2}$$

and the optimal revenues are

$$\begin{aligned} R &= \sum_{i=2}^{k} p_i \left[F(c_i) - F(c_{i-1})\right] = \sum_{i=1}^{k} \frac{i(i-1)}{2k^2} \frac{1}{k} = \frac{1}{2k^3} \sum_{i=1}^{k} \left(i^2 - i\right) \\ &= \frac{1}{2k^3} \left(\frac{k(k+1)(2k+1)}{6} - \frac{k(1+k)}{2} \right) = \frac{1}{6} \left(1 - \frac{1}{k^2}\right) \end{aligned}$$

The consumer's welfare consists of two parts: (1) the increase in the welfare due to a more efficient allocation and (2) the decrease in the welfare due to monetary transfer to the provider. The second part is equal to the revenue of the provider. The aggregated welfare from the allocation for the cutoffs $c_1, ..., c_{k-1}$ is given by

$$
\begin{aligned}
& -\int_{c_{k-1}}^{\bar{c}} \frac{1-F(c_{k-1})}{2} cf(c)\,dc - \int_{c_{k-2}}^{c_{k-1}} \left(1-F(c_{k-1}) + \frac{F(c_{k-1})-F(c_{k-2})}{2}\right) cf(c)\,dc \\
& -\int_{c_{k-3}}^{c_{k-2}} \left(1-F(c_{k-2}) + \frac{F(c_{k-2})-F(c_{k-3})}{2}\right) cf(c)\,dc \\
& -... - \int_{0}^{c_1} \left(1-F(c_1) + \frac{F(c_1)-F(c_0)}{2}\right) cf(c)\,dc \\
= & -\int_{c_{k-1}}^{\bar{c}} \left(1 - \frac{F(\bar{c})+F(c_{k-1})}{2}\right) cf(c)\,dc - \int_{c_{k-2}}^{c_{k-1}} \left(1 - \frac{F(c_{k-1})+F(c_{k-2})}{2}\right) cf(c)\,dc \\
& -\int_{c_{k-3}}^{c_{k-2}} \left(1 - \frac{F(c_{k-2})+F(c_{k-3})}{2}\right) cf(c)\,dc - ... - \int_{0}^{c_1} \left(1 - \frac{F(c_1)+F(c_0)}{2}\right) cf(c)\,dc \\
= & -\int_{0}^{\bar{c}} cf(c)\,dc + \int_{c_{k-1}}^{\bar{c}} \frac{F(\bar{c})+F(c_{k-1})}{2} cf(c)\,dc + \int_{c_{k-2}}^{c_{k-1}} \frac{F(c_{k-1})+F(c_{k-2})}{2} cf(c)\,dc \\
& + \int_{c_{k-3}}^{c_{k-2}} \frac{F(c_{k-2})+F(c_{k-3})}{2} cf(c)\,dc + ... + \int_{0}^{c_1} \frac{F(c_1)+F(c_0)}{2} cf(c)\,dc \\
= & -\mathbb{E}(c) + \frac{1}{2}\int_{c_{k-1}}^{\bar{c}} cf(c)\,dc + \sum_{i=1}^{k-1} \frac{F(c_{k-i})}{2} \int_{c_{k-i-1}}^{c_{k-i+1}} cf(c)\,dc.
\end{aligned}
$$

Plugging the expressions of the optimal cutoffs for the uniform distribution gives us the customers' welfare from the improved allocation is

$$
-\frac{1}{2} + \frac{1}{3} - \frac{1}{12k^2}.
$$

Proof of Proposition 6. Assume that $\{c_1, c_2, ..., c_{k-1}\}$ are the profit-maximizing cutoffs in the problem with k priority classes where $c_{k-1} < \bar{c}$. In the problem with $k+1$ classes set cutoffs $\{c'_1, c'_2, ..., c'_{k-1}, c_k\}$ such that $c'_i = c_i$ for any $i \in \{1, ..., k-1\}$ and $c_k \in (c_{k-1}, \bar{c})$ (in case $c_{k-1} = \bar{c}$ choose the highest cutoff below $\bar{c}$ and split the interval between this cutoff and $\bar{c}$ into two intervals). The allocation induced by cutoffs $\{c'_1, c'_2, ..., c'_{k-1}, c_k\}$ generates higher total welfare than the allocation $\{c_1, c_2, ..., c_{k-1}\}$ as the allocation $\{c'_1, c'_2, ..., c'_{k-1}, c_k\}$ has lower overall waiting costs. The proof of Proposition 5 implies that the consumers' welfare in the allocation induced by $\{c'_1, c'_2, ..., c'_{k-1}, c_k\}$ is lower than in the one induced by $\{c_1, c_2, ..., c_{k-1}\}$. Hence, the provider's profits are higher. Optimizing over the cutoffs $\{c'_1, c'_2, ..., c'_{k-1}, c_k\}$ further increases the provider's profits. Moreover, this reoptimization over the cutoffs does not eliminate any priority classes as otherwise it would allow a further increase in the provider's profits by repeating the first argument of this proof (i.e., adding the eliminated class by splitting the interval of the highest priority class). ■

Proof of Lemma 1. Assume that in the regime without priority service the provider optimally sets a price of p for its service. First, observe that setting $\pi = \Pi = p$ (which implies $c^I = 0$) in the regime with priority service replicates the same revenue as in the regime with regular service only. The provider's expected revenue (after inserting expressions (3) and (4)) is

$$
\begin{aligned}
& R = \pi F\left(c^I\right) + \Pi\left[F\left(c^{II}\right) - F\left(c^I\right)\right] \\
= & \left(V - c^{II}\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2} - c^I\frac{F\left(c^{II}\right)}{2}\right) F\left(c^I\right) + \left(V - c^{II}\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}\right)\left[F\left(c^{II}\right) - F\left(c^I\right)\right] \\
= & VF\left(c^{II}\right) - c^{II}\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}F\left(c^I\right) - c^I\frac{F\left(c^{II}\right)F\left(c^I\right)}{2} - c^{II}\frac{\left(F\left(c^{II}\right) - F\left(c^I\right)\right)^2}{2} \\
= & VF\left(c^{II}\right) - c^{II}\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}F\left(c^{II}\right) - c^I\frac{F\left(c^{II}\right)F\left(c^I\right)}{2}.
\end{aligned}
$$

Instead of optimizing over prices (Π, π), we can optimize over the cutoffs (c^I, c^{II}). Taking derivative w.r.t. c^I gives

$$
\begin{aligned}
\frac{\partial R}{\partial c^I} &= c^{II}\frac{f\left(c^I\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^I\right)}{2} - c^I\frac{F\left(c^{II}\right)f\left(c^I\right)}{2} \\
&= \left(c^{II} - c^I\right)\frac{f\left(c^I\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^I\right)}{2}.
\end{aligned}
$$

This derivative at $c^I = 0$ equals $c^{II}\frac{f(0)}{2}F\left(c^{II}\right) > 0$, where the inequality follows since in the revenue maximizing allocation $c^{II} > 0$ and $f(c) > 0$ holds for all[32] $c \in [0, \bar{c}]$. ■

Proof of Proposition 7. We first show that under the conditions of the proposition the monopolist sets the price so that no consumers are excluded. We start with the regime in which no priority service is provided. The provider's revenues are $pF(c^r)$.

Instead of maximizing over p we can optimize over c^r, noting that

$$p = V - c^r\frac{F(c^r)}{2}.$$

Hence the provider's revenue is

$$\left(V - c^r\frac{F(c^r)}{2}\right)F(c^r).$$

The FOC is

$$Vf(c^r) - \frac{(F(c^r))^2}{2} - c^rF(c^r)f(c^r) \geq 0,$$

where the equality holds in the case of $c^r \in (0, \bar{c})$. The optimal cutoff c^r satisfies

$$V \geq c^rF(c^r) + \frac{(F(c^r))^2}{2f(c^r)}.$$

[32] If $f(0) = 0$ then assuming differentiability (or one sided differentiability) of f implies that $f'(0) > 0$ and the result follows from the second order condition.

DRFR implies that the right-hand side is monotone in c^r, and hence $f(\bar{c})[V-\bar{c}] \geq \frac{1}{2}$ is sufficient for the optimal cutoff c^r to be $\bar{c}$. Therefore, the optimal price is

$$p = V - \frac{\bar{c}}{2}.$$

and the consumers' surplus equals

$$\int_0^{\bar{c}} \left(V - p - \frac{c}{2}\right) f(c)dc = \frac{1}{2}\int_0^{\bar{c}} (\bar{c} - c) f(c)dc = \frac{\bar{c} - \mathbb{E}(c)}{2}.$$

Assume now that the provider offers both regular and priority services. The provider's expected revenue (in the second line we plug expressions (3) and (4)) is[33]

$$\begin{aligned} R &= \pi F\left(c^I\right) + \Pi\left[F\left(c^{II}\right) - F\left(c^I\right)\right] \\ &= VF\left(c^{II}\right) - c^{II}\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}F\left(c^{II}\right) - c^I\frac{F\left(c^{II}\right)F\left(c^I\right)}{2}. \end{aligned}$$

Taking the derivatives w.r.t. c^I and c^{II} gives

$$\begin{aligned} \frac{\partial R}{\partial c^{II}} &= Vf\left(c^{II}\right) - \frac{F\left(c^{II}\right) - F\left(c^I\right)}{2}F\left(c^{II}\right) - c^{II}f\left(c^{II}\right)\frac{F\left(c^{II}\right) - F\left(c^I\right)}{2} \\ &\quad - c^{II}\frac{f\left(c^{II}\right)}{2}F\left(c^{II}\right) - c^I\frac{f\left(c^{II}\right)F\left(c^I\right)}{2}; \\ \frac{\partial R}{\partial c^I} &= c^{II}\frac{f\left(c^I\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^I\right)}{2} - c^I\frac{F\left(c^{II}\right)f\left(c^I\right)}{2} \\ &= \left(c^{II} - c^I\right)\frac{f\left(c^I\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^I\right)}{2}. \end{aligned}$$

At the optimal cutoff c^{II} we have $\frac{\partial R}{\partial c^{II}} \geq 0$, where the equality holds if the optimal cutoff $c^{II} \in (0, \bar{c})$. That is.

$$V \geq c^{II}F\left(c^{II}\right) + \frac{\left(F\left(c^{II}\right)\right)^2}{2f\left(c^{II}\right)} - \frac{F\left(c^I\right)}{2}\left(\frac{F\left(c^I\right)}{f\left(c^I\right)} + \frac{F\left(c^{II}\right)}{f\left(c^{II}\right)}\right).$$

Observe that DRFR and $f(\bar{c})[V-\bar{c}] \geq \frac{1}{2}$ imply that for any $c \in (0, \bar{c})$ we have

$$V > cF(c) + \frac{(F(c))^2}{2f(c)}$$

and hence for any c^{II},$c^I \in (0, \bar{c})$ we have

$$V > c^{II}F\left(c^{II}\right) + \frac{\left(F\left(c^{II}\right)\right)^2}{2f\left(c^{II}\right)} - \frac{F\left(c^I\right)}{2}\left(\frac{F\left(c^I\right)}{f\left(c^I\right)} + \frac{F\left(c^{II}\right)}{f\left(c^{II}\right)}\right),$$

[33] For the detailed derivation of the next expression see the proof of Lemma 1.

which implies that $c^{II} = \bar{c}$. Hence the provider sets the prices Π and π such that $c^{II} = \bar{c}$. Moreover, the type c^I that is indifferent between the two options satisfies

$$V - c^I \frac{1 - F\left(c^I\right)}{2} - \Pi = V - c^I \left(1 - \frac{F\left(c^I\right)}{2}\right) - \pi.$$

The last equality that specifies the indifference type c^I implies that

$$\Pi - \pi = c^I \left(1 - \frac{F\left(c^I\right)}{2} - \frac{1 - F\left(c^I\right)}{2}\right) = \frac{c^I}{2}.$$

Moreover, type $c^{II} = \bar{c}$ is indifferent between non-participation and consuming the priority service:

$$V - \Pi - \bar{c}\frac{1 - F\left(c^I\right)}{2} = 0$$

The last two equalities imply

$$\begin{aligned} \Pi &= V - \bar{c}\frac{1 - F\left(c^I\right)}{2} \\ \pi &= \Pi - \frac{c^I}{2} = V - \bar{c}\frac{1 - F\left(c^I\right)}{2} - \frac{c^I}{2}. \end{aligned}$$

For a fixed c^I the aggregated customers' welfare is

$$\begin{aligned} & \int_0^{c^I} \left(V - c\left(1 - F\left(c^I\right) + \frac{F\left(c^I\right)}{2}\right) - \pi\right) f(c)dc + \int_{c^I}^{\bar{c}} \left(V - c\frac{1 - F\left(c^I\right)}{2} - \Pi\right) f(c)dc \\ = & \left(\bar{c}\frac{1 - F\left(c^I\right)}{2} + \frac{c^I}{2}\right) F\left(c^I\right) - \left(1 - \frac{F\left(c^I\right)}{2}\right) \int_0^{c^I} cf(c)dc \\ & + \bar{c}\frac{1 - F\left(c^I\right)}{2}\left(1 - F\left(c^I\right)\right) - \frac{1 - F\left(c^I\right)}{2} \int_{c^I}^{\bar{c}} cf(c)dc \\ = & \ \bar{c}\frac{1 - F\left(c^I\right)}{2} + \frac{c^I}{2}F\left(c^I\right) - \frac{1 - F\left(c^I\right)}{2}\mathbb{E}(c) - \frac{1}{2}\int_0^{c^I} cf(c)dc. \end{aligned}$$

We now show that if distribution F satisfies DRFR, then for any $c^I \in (0, \bar{c})$ the customers' welfare under a monopolistic regime that offers only regular service is higher than the customers' welfare if the monopolist offers both regular service and priority service. That is, it is sufficient to show that

$$\frac{\bar{c}}{2} - \frac{E(c)}{2} \geq \bar{c}\frac{1 - F\left(c^I\right)}{2} + \frac{c^I}{2}F\left(c^I\right) - \frac{1 - F\left(c^I\right)}{2}\mathbb{E}(c) - \frac{1}{2}\int_0^{c^I} cf(c)dc$$

for any $c^I \in (0, \bar{c})$. The last inequality is equivalent to

$$c^I F\left(c^I\right) - \bar{c}F\left(c^I\right) + F\left(c^I\right)\mathbb{E}(c) - \int_0^{c^I} cf(c)dc \leq 0 \tag{8}$$

Observe that for $c^I = 0$ and for $c^I = \bar{c}$, the left-hand side is equal to zero. Deriving the left-hand side of the last inequality with respect to c^I gives us

$$\begin{aligned} & F\left(c^I\right) + c^I f\left(c^I\right) - \bar{c} f\left(c^I\right) + f\left(c^I\right)\mathbb{E}(c) - c^I f\left(c^I\right) \\ = \; & F\left(c^I\right) - \bar{c} f\left(c^I\right) + f\left(c^I\right)\mathbb{E}(c) = f\left(c^I\right)\left[\frac{F\left(c^I\right)}{f\left(c^I\right)} - \bar{c} + \mathbb{E}(c)\right]. \end{aligned}$$

DRFR implies that the derivative of the left-hand side of (8) changes its sign once from negative to positive, and hence (8) holds for any $c^I \in (0, \bar{c})$. ■

Proof of Proposition 8. Start with types $c \geq c^I$. Observe that type $\bar{c}$ keeps the same utility (which is zero) in the optimal mechanism when the priority service is not offered, and in the optimal mechanism in which priority service is offered. In other words, for this type the decrease in the waiting costs exactly offsets the increase in the price $\Pi - p$ that this consumer is paying. All types in the interval $[c^I, \bar{c})$, while facing the same increase in the price, enjoy a smaller decrease in the waiting costs relative to type $\bar{c}$. Hence, these types are worse off in the regime with priority service.

Next observe that if F is first-order stochastically dominated by the uniform distribution, then for every $c \in (0, \bar{c})$ we have $\bar{c}F(c) - c \geq \bar{c}\frac{c}{\bar{c}} - c > 0$. Hence $\pi > p$, which implies that the types in the interval $[0, c^I)$ in the regime with priority service, under the assumptions of the proposition, face a higher price and a higher waiting cost. Therefore, these consumers are worse off as well. ■

Proof of Lemma 2. The revenue-maximizing cutoff c^r in the regime where only the regular service is offered satisfies the first-order conditions (see proof of Proposition 7):

$$V \geq c^r F(c^r) + \frac{\left(F(c^r)\right)^2}{2f(c^r)}, \tag{9}$$

where the equality holds in the case of $c^r \in (0, \bar{c})$. DRFR implies that the right-hand side is monotone in c^r, and hence $f(\bar{c})[V - \bar{c}] < \frac{1}{2f(\bar{c})}$ implies the optimal cutoff $c^r < \bar{c}$, and that in the optimal cutoff the last inequality (9) holds as a strict equality.

In the regime where both the regular service and the priority service are provided, the proof of Proposition 7 implies that at the optimal cutoff c^{II} we have $\frac{\partial R}{\partial c^{II}} \geq 0$ where the equality holds if the optimal cutoff is $c^{II} \in (0, \bar{c})$. That is,

$$V \geq c^{II} F\left(c^{II}\right) + \frac{\left(F\left(c^{II}\right)\right)^2}{2f\left(c^{II}\right)} - \frac{F\left(c^I\right)}{2}\left(\frac{F\left(c^I\right)}{f\left(c^I\right)} + \frac{F\left(c^{II}\right)}{f\left(c^{II}\right)}\right).$$

If in the optimal cutoff c^{II} the last expression holds as a strict inequality, then $c^{II} = \bar{c} > c^r$. If, by contrast, the optimal c^{II} satisfies the last expression as equality, then for any $c^I > 0$ (which follows from the proof of Lemma 1) we have

$$cF(c) + \frac{\left(F(c)\right)^2}{2f(c)} > cF(c) + \frac{\left(F(c)\right)^2}{2f(c)} - \frac{F\left(c^I\right)}{2}\left(\frac{F\left(c^I\right)}{f\left(c^I\right)} + \frac{F(c)}{f(c)}\right).$$

Monotonicity of $cF(c)+\frac{(F(c))^2}{2f(c)}$ (which follows from DRFR) implies that $c^r < c^{II}$. ■

Proof of Proposition 9. We start with the case where the provider offers regular service only at price p. Assume that consumers with waiting costs below c^r join the queue while the rest doesn't. Since $c^r < V < \bar{c} = 1$ we have partial coverage of the market (see Lemma 2). The utility of consumer with waiting costs c from joining the service is given by

$$V - p - c\frac{F(c^r)}{2}.$$

Since the utility from not joining is 0 the marginal type c^r should be indifferent between joining the queue and remaining unserved, that is c^r satisfies

$$V - p - c^r\frac{F(c^r)}{2} = 0.$$

The provider's revenues are $pF(c^r)$.

Instead of maximizing over p we can optimize over c^r noting that

$$p = V - c^r\frac{F(c^r)}{2}.$$

Hence the provider's profits are

$$\left(V - c^r\frac{F(c^r)}{2}\right)F(c^r)$$

The FOC is

$$Vf(c^r) - \frac{(F(c^r))^2}{2} - c^rF(c^r)f(c^r) = 0.$$

We can look at the FOC wrt c^r because we do not have a full coverage $c^r \in (0,1)$. The optimal cutoff c^r satisfies

$$V = c^rF(c^r) + \frac{(F(c^r))^2}{2f(c^r)}.$$

The consumers' surplus is equal to

$$\begin{aligned}\int_0^{c^r}\left(V - p - c\frac{F(c^r)}{2}\right)f(c)dc &= \int_0^{c^r}\left(c^r\frac{F(c^r)}{2} - c\frac{F(c^r)}{2}\right)f(c)dc \qquad (10)\\ &= \frac{F(c^r)}{2}\int_0^{c^r}(c^r - c)f(c)dc\\ &= \frac{F(c^r)}{2}\int_0^{c^r}F(c)dc\end{aligned}$$

where the last equality follows from integration by parts.

Assume now that the provider sets two prices: one for the regular service (π) and one for priority service (Π). Consumers with very low waiting costs choose regular service, consumers with very high waiting costs do not join any service, while consumers in the middle-range join priority service. The utility of consumer with waiting costs c from joining the regular service is

$$V - \pi - c\left[\frac{F\left(c^{I}\right)}{2} + F\left(c^{II}\right) - F\left(c^{I}\right)\right] = V - \pi - c\left[F\left(c^{II}\right) - \frac{F\left(c^{I}\right)}{2}\right]$$

while the utility of consumer with waiting costs c from joining priority service is

$$V - \Pi - c\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2}.$$

Type c^{I} is indifferent between getting the regular and the priority services, while type c^{II} is indifferent between getting the priority service and no service at all. That is,

$$\begin{aligned} V - \pi - c^{I}\left[F\left(c^{II}\right) - \frac{F\left(c^{I}\right)}{2}\right] &= V - \Pi - c^{I}\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2} \Leftrightarrow \\ \Pi - \pi &= c^{I}\frac{F\left(c^{II}\right)}{2} \end{aligned}$$

and

$$V - \Pi - c^{II}\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2} = 0.$$

Hence

$$\begin{aligned} \Pi &= V - c^{II}\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2} \\ \pi &= V - c^{II}\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2} - c^{I}\frac{F\left(c^{II}\right)}{2}. \end{aligned}$$

The seller's expected revenue is

$$\pi F\left(c^{I}\right) + \Pi\left[F\left(c^{II}\right) - F\left(c^{I}\right)\right]$$

while consumers' surplus is

$$\int_{0}^{c^{I}}\left(V - \pi - c\left[F\left(c^{II}\right) - \frac{F\left(c^{I}\right)}{2}\right]\right) f(c)dc + \int_{c^{I}}^{c^{II}}\left(V - \Pi - c\frac{F\left(c^{II}\right) - F\left(c^{I}\right)}{2}\right) f(c)dc.$$

Plugging in the expressions for the prices into the consumers' surplus we get

$$\int_0^{c^I}\left(V-\pi-c\left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\right)f(c)dc+\int_{c^I}^{c^{II}}\left(V-\Pi-c\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\right)f(c)dc \tag{11}$$

$$\begin{aligned}
&= \int_0^{c^I}\left(c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}+c^I\frac{F\left(c^{II}\right)}{2}-c\left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\right)f(c)dc+ \\
&\quad +\int_{c^I}^{c^{II}}\left(c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}-c\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\right)f(c)dc \\
&= \left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\int_0^{c^I}\left(c^I-c\right)f(c)dc+\int_0^{c^{II}}\left(c^{II}-c^I\right)\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}f(c)dc \\
&\quad +\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\int_{c_1}^{c^{II}}\left(c^{II}-c\right)f(c)dc \\
&= \left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\int_0^{c^I}\left(c^I-c\right)f(c)dc+\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\int_{c^I}^{c^{II}}\left(c^{II}-c\right)f(c)dc \\
&\quad +\left(c^{II}-c^I\right)\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}F\left(c^I\right) \\
&= \left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\int_0^{c^I}F(c)dc+\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\left[-\left(c^{II}-c^I\right)F\left(c^I\right)+\int_{c^I}^{c^{II}}F(c)dc\right] \\
&\quad +\left(c^{II}-c^I\right)\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}F\left(c^I\right) \\
&= \left[F\left(c^{II}\right)-\frac{F\left(c^I\right)}{2}\right]\int_0^{c^I}F(c)dc+\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\int_{c^I}^{c^{II}}F(c)dc \\
&= F\left(c^{II}\right)\int_0^{c^I}F(c)dc-\frac{F\left(c^I\right)}{2}\int_0^{c^{II}}F(c)dc+\frac{F\left(c^{II}\right)}{2}\int_{c^I}^{c^{II}}F(c)dc \\
&= \frac{F\left(c^{II}\right)}{2}\int_0^{c^{II}}F(c)dc+\frac{F\left(c^{II}\right)}{2}\int_0^{c^I}F(c)dc-\frac{F\left(c^I\right)}{2}\int_0^{c^{II}}F(c)dc.
\end{aligned}$$

We now find the optimal cutoffs. The seller's expected revenue is

$$\begin{aligned}
& R=\pi F\left(c^I\right)+\Pi\left[F\left(c^{II}\right)-F\left(c^I\right)\right] \\
&= \left(V-c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}-c^I\frac{F\left(c^{II}\right)}{2}\right)F\left(c^I\right)+\left(V-c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}\right)\left[F\left(c^{II}\right)-F\left(c^I\right)\right] \\
&= VF\left(c^{II}\right)-c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}F\left(c^I\right)-c^I\frac{F\left(c^{II}\right)F\left(c^I\right)}{2}-c^{II}\frac{\left(F\left(c^{II}\right)-F\left(c^I\right)\right)^2}{2} \\
&= VF\left(c^{II}\right)-c^{II}\frac{F\left(c^{II}\right)-F\left(c^I\right)}{2}F\left(c^{II}\right)-c^I\frac{F\left(c^{II}\right)F\left(c^I\right)}{2}.
\end{aligned}$$

Maximizing it w.r.t. c^I and c^{II} gives the FOC

$$\begin{aligned}\frac{\partial R}{\partial c^{II}} &= Vf\left(c^{II}\right) - \frac{F\left(c^{II}\right)-F\left(c^{I}\right)}{2}F\left(c^{II}\right) - c^{II}f\left(c^{II}\right)\frac{F\left(c^{II}\right)-F\left(c^{I}\right)}{2} \\ &\quad -c^{II}\frac{f\left(c^{II}\right)}{2}F\left(c^{II}\right) - c^{I}\frac{f\left(c^{II}\right)F\left(c^{I}\right)}{2} = 0 \\ \frac{\partial R}{\partial c^{I}} &= c^{II}\frac{f\left(c^{I}\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^{I}\right)}{2} - c^{I}\frac{F\left(c^{II}\right)f\left(c^{I}\right)}{2} \\ &= \left(c^{II}-c^{I}\right)\frac{f\left(c^{I}\right)}{2}F\left(c^{II}\right) - \frac{F\left(c^{II}\right)F\left(c^{I}\right)}{2} = 0\end{aligned}$$

Since $c^{II}=0$ is not a maximum we get

$$\frac{\partial R}{\partial c^{I}} = 0 \Longleftrightarrow \left(c^{II}-c^{I}\right)f\left(c^{I}\right) - F\left(c^{I}\right) = 0 \Longleftrightarrow c^{II} = c^{I} + \frac{F\left(c^{I}\right)}{f\left(c^{I}\right)}$$

and $\frac{\partial R}{\partial c^{II}} = 0 \Longleftrightarrow$

$$V - \frac{F\left(c^{II}\right)-F\left(c^{I}\right)}{2}\frac{F\left(c^{II}\right)}{f\left(c^{II}\right)} - c^{II}\left[F\left(c^{II}\right) - \frac{F\left(c^{I}\right)}{2}\right] - c^{I}\frac{F\left(c^{I}\right)}{2} = 0$$

Alternative writing $\frac{\partial R}{\partial c^{II}} = 0 \Longleftrightarrow$

$$V - \frac{F\left(c^{II}\right)-F\left(c^{I}\right)}{2}\left[\frac{F\left(c^{II}\right)}{f\left(c^{II}\right)} + c^{II}\right] - c^{II}\frac{F\left(c^{II}\right)}{2} - c^{I}\frac{F\left(c^{I}\right)}{2} = 0.$$

Assume that $F(c) = c^{\theta}$ for $c \in [0,1]$ and $f(c) = \theta c^{\theta-1}$. Then **without priority service** we have

$$\begin{aligned}V &= c^{r}F\left(c^{r}\right) + \frac{\left(F\left(c^{r}\right)\right)^{2}}{2f\left(c^{r}\right)} \Longleftrightarrow \\ V &= c^{r}\left(c^{r}\right)^{\theta} + \frac{\left(c^{r}\right)^{2\theta}}{2\theta\left(c^{r}\right)^{\theta-1}} \Longleftrightarrow V = \left(c^{r}\right)^{\theta+1} + \frac{\left(c^{r}\right)^{\theta+1}}{2\theta} \\ V &= \left(c^{r}\right)^{\theta+1}\left(1+\frac{1}{2\theta}\right) \Longleftrightarrow \left(c^{r}\right)^{\theta+1} = \frac{2\theta V}{2\theta+1} \Longleftrightarrow c^{r} = \left(\frac{2\theta V}{2\theta+1}\right)^{\frac{1}{\theta+1}}\end{aligned}$$

Plugging the expression for c^r into the expression for consumers' surplus given in (10) gives

$$\frac{F\left(c^{r}\right)}{2}\int_{0}^{c^{r}}F(c)dc = \frac{\left(c^{r}\right)^{\theta}}{2}\int_{0}^{c^{r}}c^{\theta}dc = \frac{1}{2}\frac{1}{\theta+1}\left(c^{r}\right)^{2\theta+1} = \frac{1}{2}\frac{1}{\theta+1}\left(\frac{2\theta V}{2\theta+1}\right)^{\frac{2\theta+1}{\theta+1}}. \quad (12)$$

While the price is

$$p = V - c^{r}\frac{F\left(c^{r}\right)}{2} = V - \frac{1}{2}\left(\frac{2\theta V}{2\theta+1}\right)^{\frac{1}{\theta+1}}\left(\frac{2\theta V}{2\theta+1}\right)^{\frac{\theta}{\theta+1}} = V - \frac{1}{2}\left(\frac{2\theta V}{2\theta+1}\right).$$

With priority service we get

$$c^{II} = c^I + \frac{c^I}{\theta} \iff c^{II} = c^I \frac{\theta+1}{\theta}$$

and

$$\begin{aligned}
V - \frac{\left(c^{II}\right)^\theta - \left(c^I\right)^\theta}{2}\left[c^{II} + \frac{c^{II}}{\theta}\right] - \frac{\left(c^{II}\right)^{\theta+1}}{2} - \frac{\left(c^I\right)^{\theta+1}}{2} &= 0 \iff \\
V - \frac{\left(c^{II}\right)^\theta - \left(c^I\right)^\theta}{2} c^{II}\frac{\theta+1}{\theta} - \frac{\left(c^{II}\right)^{\theta+1}}{2} - \frac{\left(c^I\right)^{\theta+1}}{2} &= 0 \iff \\
V - \frac{\left(c^{II}\right)^{\theta+1}}{2}\frac{2\theta+1}{\theta} + \frac{\left(c^I\right)^\theta}{2} c^{II}\frac{\theta+1}{\theta} - \frac{\left(c^I\right)^{\theta+1}}{2} &= 0
\end{aligned}$$

Plugging the expression we got for c^{II} the last equation becomes

$$V - \frac{\left(c^I\right)^{\theta+1}}{2}\left(\frac{\theta+1}{\theta}\right)^{\theta+1}\frac{2\theta+1}{\theta} + \frac{\left(c^I\right)^{\theta+1}}{2}\left(\frac{\theta+1}{\theta}\right)^2 - \frac{\left(c^I\right)^{\theta+1}}{2} = 0 \iff$$

$$V = \frac{\left(c^I\right)^{\theta+1}}{2}\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1}\frac{2\theta+1}{\theta} - \left(\frac{\theta+1}{\theta}\right)^2 + 1\right] \iff$$

$$2V = \left(c^I\right)^{\theta+1}\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1}\frac{2\theta+1}{\theta} - \frac{2\theta+1}{\theta^2}\right] \iff$$

$$2V = \left(c^I\right)^{\theta+1}\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1} - \frac{1}{\theta}\right]\frac{2\theta+1}{\theta} \iff$$

$$\frac{2\theta V}{(2\theta+1)\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1} - \frac{1}{\theta}\right]} = \left(c^I\right)^{\theta+1} \iff$$

$$c^I = \left(\frac{2\theta V}{(2\theta+1)\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1} - \frac{1}{\theta}\right]}\right)^{\frac{1}{\theta+1}}$$

$$c^{II} = \frac{\theta+1}{\theta}\left(\frac{2\theta V}{(2\theta+1)\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1} - \frac{1}{\theta}\right]}\right)^{\frac{1}{\theta+1}}$$

Plugging the expressions for c^I and c^{II} into the expression for consumers' surplus given in (11) gives

$$\frac{F\left(c^{II}\right)}{2}\int_0^{c^{II}}F(c)dc+\frac{F\left(c^{II}\right)}{2}\int_0^{c^{I}}F(c)dc-\frac{F\left(c^{I}\right)}{2}\int_0^{c^{II}}F(c)dc \tag{13}$$

$$\begin{aligned}
&= \frac{1}{2}\left(\left(c^{II}\right)^{\theta}\int_0^{c^{II}}(c)^{\theta}dc+\left(c^{II}\right)^{\theta}\int_0^{c^{I}}(c)^{\theta}dc-\left(c^{I}\right)^{\theta}\int_0^{c^{II}}(c)^{\theta}dc\right)\\
&= \frac{1}{2}\frac{1}{\theta+1}\left(\left(c^{II}\right)^{2\theta+1}+\left(c^{II}\right)^{\theta}\left(c^{I}\right)^{\theta+1}-\left(c^{I}\right)^{\theta}\left(c^{II}\right)^{\theta+1}\right)\\
&= \frac{1}{2}\frac{1}{\theta+1}\left(c^{I}\right)^{2\theta+1}\left(\left(\frac{\theta+1}{\theta}\right)^{2\theta+1}+\left(\frac{\theta+1}{\theta}\right)^{\theta}-\left(\frac{\theta+1}{\theta}\right)^{\theta+1}\right)\\
&= \frac{1}{2}\frac{1}{\theta+1}\left(\frac{2\theta V}{(2\theta+1)\left[\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right]}\right)^{\frac{2\theta+1}{\theta+1}}\left(\frac{\theta+1}{\theta}\right)^{\theta}\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)\\
&= \left(\frac{2\theta V}{(2\theta+1)}\right)^{\frac{2\theta+1}{\theta+1}}\frac{\left(\frac{\theta+1}{\theta}\right)^{\theta-1}\frac{1}{2\theta}}{\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}}.
\end{aligned}$$

We show now that the consumers' surplus if no priority service is offered (12) exceeds the one if it is offered (13):We will show now that

$$\left(\frac{2\theta V}{(2\theta+1)}\right)^{\frac{2\theta+1}{\theta+1}}\frac{\left(\frac{\theta+1}{\theta}\right)^{\theta-1}\frac{1}{2\theta}}{\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}}>\frac{1}{2}\frac{1}{\theta+1}\left(\frac{2\theta V}{2\theta+1}\right)^{\frac{2\theta+1}{\theta+1}}\iff$$

$$\frac{\left(\frac{\theta+1}{\theta}\right)^{\theta-1}\frac{1}{2\theta}}{\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}}>\frac{1}{2}\frac{1}{\theta+1}\iff$$

$$\frac{\left(\frac{\theta+1}{\theta}\right)^{\theta-1}\frac{\theta+1}{\theta}}{\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}}>1\iff$$

$$\left(\frac{\theta+1}{\theta}\right)^{\theta}>\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}\iff$$

$$\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}\right)^{\frac{\theta}{\theta+1}}>\left(\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}\right)^{\frac{\theta}{\theta+1}}\iff$$

$$\left(\frac{\theta+1}{\theta}\right)^{\theta+1}>\left(\frac{\theta+1}{\theta}\right)^{\theta+1}-\frac{1}{\theta}$$

which holds for any $\theta > 0$. ■

Proof of Lemma 3. Observe that

$$B'(s) = \frac{-g(s)}{1-s} + \frac{\int_s^1 g(t)\,dt}{(1-s)^2} - \frac{g(s)}{s} + \frac{\int_0^s g(t)\,dt}{s^2}.$$

Assume that g is concave. Since g is differentiable, the mean value theorem implies that for any $t < s$ there exists $z(t) \in (t,s)$ such that $g(t) = g(s) - g'(z(t))(s-t)$. Similarly, for any $t > s$ there exists $z(t) \in (s,t)$ such that $g(t) = g(s) + g'(z(t))(t-s)$. Plugging the expressions for $g(t)$ gives

$$\begin{aligned} B'(s) &= \frac{-g(s)}{1-s} + \frac{\int_s^1 (g(s) + g'(z(t))(t-s))\,dt}{(1-s)^2} - \frac{g(s)}{s} + \frac{\int_0^s (g(s) - g'(z(t))(s-t))\,dt}{s^2} \\ &= \frac{\int_s^1 g'(z(t))(t-s)dt}{(1-s)^2} - \frac{\int_0^s g'(z(t))(s-t)dt}{s^2} < \frac{\int_s^1 g'(s)(t-s)dt}{(1-s)^2} - \frac{\int_0^s g'(s)(s-t)dt}{s^2} \\ &= g'(s)\left[\frac{\int_s^1 (t-s)dt}{(1-s)^2} - \frac{\int_0^s (s-t)dt}{s^2}\right] = 0, \end{aligned}$$

where the inequality follows from the concavity of g, which implies that g' is decreasing. For convex g we get the opposite inequality. ■

Proof of Proposition 10. We start with concave waiting costs $g(t)$. The providers' profits are $sp(s)$, where $p(s)$ is given by (5). Therefore, the optimal share is

$$\max_s s\left(\frac{\int_s^1 g(t)\,dt}{1-s} - \frac{\int_0^s g(t)\,dt}{s}\right).$$

The derivative with respect to s is

$$\begin{aligned} &\frac{\int_s^1 g(t)\,dt}{1-s} - \frac{\int_0^s g(t)\,dt}{s} + s\left(-\frac{g(s)}{1-s} + \frac{\int_s^1 g(t)\,dt}{(1-s)^2} - \frac{g(s)}{s} + \frac{\int_0^s g(t)\,dt}{s^2}\right) \\ &= \frac{\int_s^1 g(t)\,dt}{1-s}\left(1 + \frac{1}{1-s}\right) - g(s)\left(1 + \frac{1}{1-s}\right) = \frac{1}{(1-s)^2}\left(\int_s^1 g(t)\,dt - g(s)(1-s)\right) \\ &= \frac{1}{(1-s)^2}\int_s^1 (g(t) - g(s))\,dt > 0, \end{aligned}$$

where the inequality follows from monotonicity of g. Therefore, the optimal share is $s^* = 1$. Hence, the optimal price is

$$p^* = p(1) = \lim_{s=1}\left(\frac{\int_s^1 g(t)\,dt}{1-s} - \frac{\int_0^s g(t)\,dt}{s}\right) = \lim_{s=1}\frac{-g(s)}{-1} - \int_0^1 g(t)\,dt = g(1) - \int_0^1 g(t)\,dt.$$

Assume now convex g. Since Lemma 3 implies that $B(s)$ is monotone, any price $p \in [B(0), B(1)]$ will lead all customers to join the priority service. Therefore, the optimal price is $p^* = B(1)$. ■

Proof of Proposition 11.

Similarly to the linear case we will show that for any price of the priority service, and not necessarily the optimal one, the customers' welfare if priority service is offered is lower than if this service is not offered. If the provider sets the price that induces cutoff $c^* \in [0, \bar{c}]$ that divides the customers into two categories, the total customers' welfare is

$$\begin{aligned} & -\int_0^{c^*} c\frac{\int_{1-F(c^*)}^1 t^\theta dt}{F(c^*)} f(c)dc - \int_{c^*}^{\bar{c}} \left(p + c\frac{\int_0^{1-F(c^*)} t^\theta dt}{1-F(c^*)}\right) f(c)dc \\ = & -\frac{1-(1-F(c^*))^{\theta+1}}{(\theta+1)F(c^*)}\int_0^{c^*} cf(c)dc - c^*\frac{1-F(c^*)}{F(c^*)}\frac{1-(1-F(c^*))^\theta}{\theta+1} - \frac{(1-F(c^*))^\theta}{\theta+1}\int_{c^*}^{\bar{c}} cf(c)dc. \end{aligned}$$

The customers' welfare if no priority service is offered is

$$-E(c)\int_0^1 t^\theta dt = -\frac{E(c)}{\theta+1}.$$

We show that

$$-\frac{1-(1-F(c^*))^{\theta+1}}{(\theta+1)F(c^*)}\int_0^{c^*} cf(c)dc - c^*\frac{1-F(c^*)}{F(c^*)}\frac{1-(1-F(c^*))^\theta}{\theta+1} - \frac{(1-F(c^*))^\theta}{\theta+1}\int_{c^*}^{\bar{c}} cf(c)dc \le -\frac{E(c)}{\theta+1} \tag{14}$$

for any $c^* \in [0, \bar{c}]$ and $\theta \ge 1$. Rearranging (14) gives

$$\begin{aligned} & \left(-\frac{1-(1-F(c^*))^{\theta+1}}{F(c^*)} + (1-F(c^*))^\theta\right)\int_0^{c^*} cf(c)dc - c^*\frac{1-F(c^*)}{F(c^*)}\left(1-(1-F(c^*))^\theta\right) \\ & \qquad - (1-F(c^*))^\theta \int_0^{\bar{c}} cf(c)dc \le -E(c). \end{aligned}$$

We can rewrite the last inequality as

$$\begin{aligned} & -\frac{1-(1-F(c^*))^{\theta+1} - (1-F(c^*))^\theta F(c^*)}{F(c^*)}\int_0^{c^*} cf(c)dc - \\ & c^*\frac{1-F(c^*)}{F(c^*)}\left(1-(1-F(c^*))^\theta\right) \le -\left(1-(1-F(c^*))^\theta\right)E(c) \end{aligned}$$

$$\begin{aligned} -\frac{1-(1-F(c^*))^\theta}{F(c^*)}\int_0^{c^*} cf(c)dc - c^*\frac{1-F(c^*)}{F(c^*)}\left(1-(1-F(c^*))^\theta\right) & \le -\left(1-(1-F(c^*))^\theta\right)E(c) \\ -\frac{\int_0^{c^*} cf(c)dc}{F(c^*)}\int_0^{c^*} cf(c)dc - c^*\frac{1-F(c^*)}{F(c^*)} & \le -E(c), \end{aligned}$$

which is independent of θ. Since we showed the inequality for the linear case ($\theta = 1$) in Proposition 2, this inequality holds for any $\theta > 0$. ■

10.2 Appendix B. Duopoly

Proof of Proposition 12. To characterize the equilibrium of the subgame following the price announcement (p_1, p_2) we will consider a few possible profiles.

Profile (A) $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} > 0$.
Profile (B) $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} = 0$, $n_2^{np} > 0$.
Profile (C) $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} = 0$, which is symmetric to profile B.
Profile (D) $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} = 0$, $n_2^{np} = 0$.
Profile (E) $n_1^p > 0$, $n_2^p = 0$, $n_1^{np} > 0$, $n_2^{np} > 0$.
Profile (F) $n_1^p = 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} > 0$, which is symmetric to profile F.
Profile (G) $n_1^p = 0$, $n_2^p = 0$, $n_1^{np} > 0$, $n_2^{np} > 0$.
Profile (H) $n_1^p > 0$, $n_2^p = 0$, $n_1^{np} = 0$, $n_2^{np} > 0$.
Profile (I) $n_1^p = 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} = 0$, which is symmetric to profile H.

There is no equilibrium in which $n_1^p > 0$, $n_2^p = 0$, $n_1^{np} > 0$, $n_2^{np} = 0$ as customers from the regular service of provider 1 should switch to the regular service of provider 2. For a similar reason there is no equilibrium in which $n_1^p = 0$, $n_2^p > 0$, $n_1^{np} = 0$, $n_2^{np} > 0$. Also there is no equilibrium in which all the customers are concentrated at a single provider in either of the services.

We now consider each of the above profiles separately.

Profile A $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} > 0$ and $n_2^{np} > 0$. Since for both providers both classes are nonempty, in this profile the customers are indifferent between all their opportunities (provider 1 vs. provider 2, priority vs. regular services) and the equilibrium conditions are

1. $p_1 = \frac{n_1^p + n_1^{np}}{2}$
2. $p_2 = \frac{n_2^p + n_2^{np}}{2}$
3. $-p_1 - \frac{n_1^p}{2} = -p_2 - \frac{n_2^p}{2}$
4. $n_1^p + n_2^p + n_1^{np} + n_2^{np} = 1$

Observe that (1), (2), and (4) imply that $p_1 + p_2 = \frac{1}{2}$. Further we get $n_2^p + n_2^{np} = 2p_2$ and $n_1^p + n_1^{np} = 2p_1$ and $n_1^p - n_2^p = 2(p_2 - p_1)$, which implies that $n_1^p = 2 - n_2^{np} - 6p_1, n_1^{np} = 8p_1 + n_2^{np} - 2, n_2^p = 1 - n_2^{np} - 2p_1$.

Further observe that $n_1^p = 2 - n_2^{np} - 6p_1 > 0$ implies that $p_1 < \frac{1}{3}$. Symmetry implies that $p_2 < \frac{1}{3}$. In this case we have a continuum of equilibria. Conditions (1)–(3) imply that

$$\begin{aligned} n_1^p &= 2 - n_2^{np} - 6p_1 \\ n_1^{np} &= 8p_1 + n_2^{np} - 2 \\ n_2^p &= 1 - n_2^{np} - 2p_1, \end{aligned}$$

and so for any $\max\{2-8p_1,0\} < n_2^{np} < \min\{2-6p_1, 1-2p_1\}$ we have an equilibrium. Observe that provider 1 is interested in the lowest possible n_2^{np}. For future derivations, observe that for provider 1, in the best equilibrium n_1^p is (strictly) lower than $2p_1 = 1-2p_2$ if $p_1 < 1/4$ and n_1^p is (strictly) lower than $2-6p_1 = 6p_2-1$ if $p_1 > 1/4$, and the revenues are smaller than $\left(\frac{1}{2}-p_2\right)(1-2p_2)$ if $p_1 < 1/4$ (or $p_2 > 1/4$) and smaller than $(6p_2-1)\left(\frac{1}{2}-p_2\right)$ if $p_1 > 1/4$ (or $p_2 < 1/4$).

Profile B. Consider the profile with $n_1^p > 0$, $n_2^p > 0$, $n_2^{np} > 0$ and $n_1^{np} = 0$. This profile implies that $n_2 = 2p_2$. For this profile to be part of an equilibrium customers must be indifferent between getting priority service from provider 1 or 2 and regular service from provider 2. These indifference conditions imply

$$\begin{aligned} -p_1 - \frac{n_1^p}{2} &= -p_2 - \frac{n_2^p}{2} \\ p_2 &= \frac{n_2^p + n_2^{np}}{2} \\ 1 &= n_1^p + n_2^p + n_2^{np}. \end{aligned}$$

We have

$$p_2 = \frac{1-n_1^p}{2} \iff -p_1 - \frac{n_1^p}{2} = -\frac{1-n_1^p}{2} - \frac{n_2^p}{2} \iff n_1^p = \frac{1}{2} - p_1 + \frac{n_2^p}{2}.$$

Therefore,

$$p_2 = \frac{n_2^p + n_2^{np}}{2} \iff n_2^p + n_2^{np} = 2p_2 \iff n_2^{np} = 2p_2 - n_2^p$$

$$\frac{1}{2} - p_1 + \frac{n_2^p}{2} + n_2^p + 2p_2 - n_2^p = 1 \iff \frac{n_2^p}{2} = \frac{1}{2} + p_1 - 2p_2 \iff n_2^p = 1 + 2p_1 - 4p_2.$$

This implies that

$$\begin{aligned} n_2^{np} &= 2p_2 - n_2^p = 6p_2 - 1 - 2p_1 \\ n_1^p &= \frac{1}{2} - p_1 + \frac{n_2^p}{2} = 1 - 2p_2 > 0. \end{aligned}$$

In addition, for this profile to be an equilibrium, the utility from joining the regular service of provider 1 must be lower than all other options:

$$p_1 \leq \frac{n_1^p}{2} \iff 1 - 2p_2 \geq 2p_1.$$

To summarize, this profile is an equilibrium if

$$\begin{aligned} 1 + 2p_1 - 4p_2 &\geq 0 \iff p_2 \leq \frac{1}{4} + \frac{1}{2}p_1 \\ 1 - 2p_2 - 2p_1 &\geq 0 \iff p_2 \leq \frac{1}{2} - p_1 \\ 6p_2 - 1 - 2p_1 &\geq 0 \iff p_2 \geq \frac{1}{6} + \frac{1}{3}p_1. \end{aligned}$$

Profile C $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} = 0$. An analogous (to profile B) argument implies

$$\begin{aligned} n_1^p &= 1 + 2p_2 - 4p_1 \\ n_1^{np} &= 6p_1 - 1 - 2p_2 \\ n_2^p &= 1 - 2p_1 \end{aligned}$$

and this profile is part of an equilibrium if

$$\begin{aligned} 1 + 2p_2 - 4p_1 &\geq 0 \Longleftrightarrow p_1 \leq \frac{1}{4} + \frac{1}{2}p_2 \\ 1 - 2p_1 - 2p_2 &\geq 0 \Longleftrightarrow p_1 \leq \frac{1}{2} - p_2 \\ 6p_1 - 1 - 2p_2 &\geq 0 \Longleftrightarrow p_1 \geq \frac{1}{6} + \frac{1}{3}p_2. \end{aligned}$$

Profile D $n_1^p > 0$, $n_2^p > 0$, $n_1^{np} = 0$, $n_2^{np} = 0$. For this profile to be an equilibrium it must be that

1. $p_1 \leq n_1^p/2$
2. $p_2 \leq n_2^p/2$
3. $-p_1 - \frac{n_1^p}{2} = -p_2 - \frac{n_2^p}{2}$
4. $n_1^p + n_2^p = 1$

Conditions (3)+(4) imply that

$$\begin{aligned} n_1^p &= \frac{1}{2} + p_2 - p_1 \\ n_2^p &= \frac{1}{2} + p_1 - p_2, \end{aligned}$$

where $p_1 \leq n_1^p/2$ implies that $p_2 \geq 3p_1 - 1/2$ and $p_2 \leq n_2^p/2$ implies that $p_1 \geq 3p_2 - 1/2$.

Profile E $n_1^p > 0$, $n_2^p = 0$, $n_1^{np} > 0$, $n_2^{np} > 0$. For this profile to be part of an equilibrium customers must be indifferent between the priority service of provider 1, the regular service of provider 1 and the regular service of provider 2. That is

$$p_1 = \frac{n_1^p + n_1^{np}}{2} \ (E1)$$

$$-p_1 - \frac{n_1^p}{2} = -\frac{n_2^{np}}{2} \ (E2)$$

$$n_1^p + n_1^{np} + n_2^{np} = 1 \ (E3)$$

$$p_2 \geq \frac{n_2^{np}}{2} \ (E4).$$

$E1$ implies that $n_1^p + n_1^{np} = 2p_1$. Therefore, from $E3$ we get $n_2^{np} = 1 - 2p_1$. From $E2$ we get $\frac{n_1^p}{2} = \frac{n_2^{np}}{2} - p_1 = \frac{1}{2} - 2p_1 \iff n_1^p = 1 - 4p_1$. Therefore, $n_1^{np} = 2p_1 - n_1^p = 2p_1 - 1 + 4p_1 = 6p_1 - 1$. Putting the conditions together we get

$$\begin{aligned} n_1^p &= 1 - 4p_1 > 0 \iff p_1 < \frac{1}{4} \\ n_1^{np} &= 6p_1 - 1 > 0 \iff p_1 > \frac{1}{6} \\ n_2^{np} &= 1 - 2p_1 > 0 \\ p_2 &\geq \frac{1}{2} - p_1. \end{aligned}$$

Profile F $n_1^p = 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} > 0$. An analogous argument (to profile E) implies that

$$\begin{aligned} n_2^p &= 1 - 4p_2 > 0 \iff p_2 < \frac{1}{4} \\ n_2^{np} &= 6p_2 - 1 > 0 \iff p_2 > \frac{1}{6} \\ n_1^{np} &= 1 - 2p_2 > 0 \\ p_1 &\geq \frac{1}{2} - p_2. \end{aligned}$$

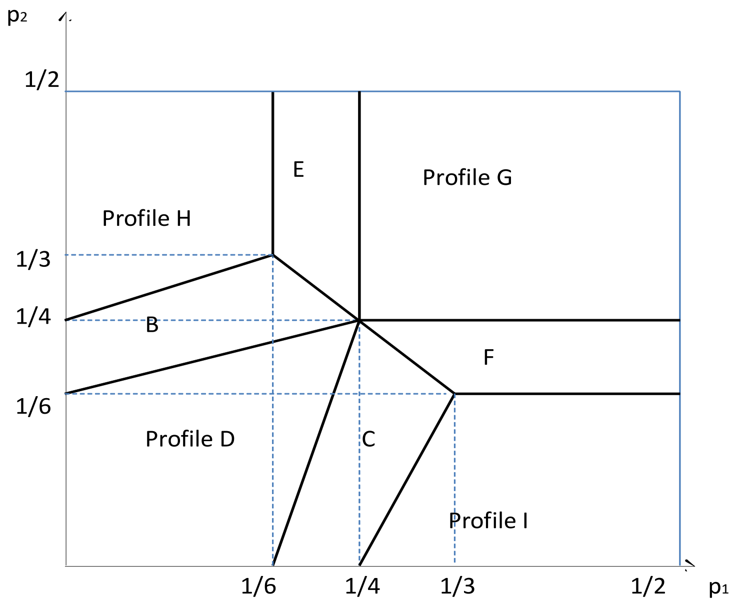
Profile G $n_1^p = 0$, $n_2^p = 0$, $n_1^{np} > 0$, $n_2^{np} > 0$. Equilibrium conditions are

$$\begin{aligned} -\frac{n_1^{np}}{2} &= -\frac{n_2^{np}}{2} \\ n_1^{np} + n_2^{np} &= 1 \\ p_1 &\geq \frac{n_1^{np}}{2} \\ p_2 &\geq \frac{n_2^{np}}{2}. \end{aligned}$$

These conditions implies that $n_1^{np} = n_2^{np} = \frac{1}{2}$, $p_1 \geq \frac{1}{4}$ and $p_2 \geq \frac{1}{4}$.

Profile H $n_1^p > 0$, $n_2^p = 0$, $n_1^{np} = 0$, $n_2^{np} > 0$. For this profile to be an equilibrium it must be satisfy

$$\begin{aligned} -p_1 - \frac{n_1^p}{2} &= -\frac{n_2^{np}}{2} \\ n_1^p + n_2^{np} &= 1 \\ p_1 &\leq \frac{n_1^p}{2} \\ p_2 &\geq \frac{n_2^{np}}{2}. \end{aligned}$$



The first two equalities imply that

$$n_1^p = \frac{1}{2} - p_1 > 0$$
$$n_2^{np} = \frac{1}{2} + p_1 > 0.$$

The conditions are $p_1 \leq \frac{n_1^p}{2} \iff p_1 \leq \frac{1}{6}$ and $p_2 \geq \frac{n_2^{np}}{2} \iff p_2 \geq \frac{1}{4} + \frac{1}{2}p_1$.

Profile I $n_1^p = 0$, $n_2^p > 0$, $n_1^{np} > 0$, $n_2^{np} = 0$. An analogous argument (to profile H) implies that in equilibrium

$$n_2^p = \frac{1}{2} - p_2 > 0$$
$$n_1^{np} = \frac{1}{2} + p_2 > 0,$$

and the conditions for this profile are $p_2 \leq \frac{1}{6}$ and $p_1 \geq \frac{1}{4} + \frac{1}{2}p_2$.

We can now plot all these profiles (profile A is the plotted diagonal line).

Now after calculating the equilibrium at the second stage (following price announcement of the providers) we can calculate the best responses of the firms at the first stage. We have to consider a few cases.

Case 1. $p_2 < 1/6$. In this case if $p_1 \leq \frac{1}{3}p_2 + \frac{1}{6}$, then we are in profile D and $n_1^p = \frac{1}{2} + p_2 - p_1$. Provider 1's maximization problem is

$$\max_{p_1} R = p_1 \left(\frac{1}{2} + p_2 - p_1 \right)$$
$$s.t. p_1 \leq \frac{1}{3}p_2 + \frac{1}{6}.$$

Observe that R is concave in p_1 and reaches its maximum at $\frac{1}{2}p_2 + \frac{1}{4} > \frac{1}{3}p_2 + \frac{1}{6}$. Therefore, the optimal $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$. If $p_1 > \frac{1}{3}p_2 + \frac{1}{6}$ and $p_1 \leq \frac{1}{4} + \frac{1}{2}p_2$, then we are in profile C and $n_1^p = 1 + 2p_2 - 4p_1$. Provider 1's maximization problem is

$$\begin{aligned} \max R &= p_1 (1 + 2p_2 - 4p_1) \\ s.t. p_1 &> \frac{1}{3}p_2 + \frac{1}{6}. \end{aligned}$$

Again, R is concave in p_1 and reaches its maximum at $\frac{1}{4}p_2 + \frac{1}{8} < \frac{1}{3}p_2 + \frac{1}{6}$. Therefore, the optimal price is $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$. If $p_1 > \frac{1}{4} + \frac{1}{2}p_2$ then we are in profile I where $n_1^p = 0$, and we can conclude that for any $p_2 < 1/6$, provider 1's best response is $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$.

Case 2. $1/4 > p_2 \geq 1/6$. If $p_1 \leq 3p_2 - \frac{1}{2}$, then we are in profile B and $n_1^p = 1 - 2p_2$ and $\frac{\partial R}{\partial p_1} = 1 - 2p_2 > 0$, and the optimal price is $p_1 = 3p_2 - \frac{1}{2}$. If $p_1 > 3p_2 - \frac{1}{2}$ and $p_1 < \frac{1}{2} - p_2$ then we know that the optimal price is $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$ (which is the same as the optimal price in profiles D and C, as was shown in case 1). For $p_1 > \frac{1}{2} - p_2$ we are in profile F where $n_1^p = 0$. Therefore, we need to compare the revenue from $p_1 = 3p_2 - \frac{1}{2}$ and $n_1^p = 1 - 2p_2$ (which is $\left(3p_2 - \frac{1}{2}\right)(1 - 2p_2)$) with the revenue from $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$ and $n_1^p = \frac{1}{2} + p_2 - p_1 = \frac{1}{2} + p_2 - \frac{1}{3}p_2 - \frac{1}{6} = \frac{1}{3} + \frac{2}{3}p_2$ (which is $\left(\frac{1}{3}p_2 + \frac{1}{6}\right)\left(\frac{1}{3} + \frac{2}{3}p_2\right)$). Therefore, for all $p_2 \leq 1/4$ we have $\left(\frac{1}{3}p_2 + \frac{1}{6}\right)\left(\frac{1}{3} + \frac{2}{3}p_2\right) > \left(3p_2 - \frac{1}{2}\right)(1 - 2p_2)$. Finally, we need to compare the revenues from this equilibrium with the revenues from the best equilibrium in profile A, in which the revenues are bounded by $(6p_2 - 1)\left(\frac{1}{2} - p_2\right)$. Since $\left(\frac{1}{2} - p_2\right)(6p_2 - 1) < \left(\frac{1}{3}p_2 + \frac{1}{6}\right)\left(\frac{1}{3} + \frac{2}{3}p_2\right)$ the best response also here is $p_1 = \frac{1}{3}p_2 + \frac{1}{6}$.

Case 3. $1/3 > p_2 \geq 1/4$. If $p_1 \leq 2p_2 - \frac{1}{2}$, then we are in profile H where $n_1^p = \frac{1}{2} - p_1$. Provider 1's maximization problem is

$$\begin{aligned} \max R &= p_1 \left(\frac{1}{2} - p_1\right) \\ s.t. p_1 &\leq 2p_2 - \frac{1}{2}. \end{aligned}$$

Again, R is concave in p_1 and reaches its maximum at $p_1 = \frac{1}{4}$. Since $\frac{1}{4} > 2p_2 - \frac{1}{2}$, the optimal price in this region is $p_1 = 2p_2 - \frac{1}{2}$, and the revenues are $\left(2p_2 - \frac{1}{2}\right)(1 - 2p_2)$. If $\frac{1}{2} - p_2 > p_1 > 2p_2 - \frac{1}{2}$, then we are in profile B and $n_1^p = 1 - 2p_2$ and $\frac{\partial R}{\partial p_1} = 1 - 2p_2 > 0$, and the optimal price is $p_1 = \frac{1}{2} - p_2$ and the revenues are $\left(\frac{1}{2} - p_2\right)(1 - 2p_2)$. If $\frac{1}{4} > p_1 > \frac{1}{2} - p_2$, then we are in profile E with $n_1^p = 1 - 4p_1$. In profile E provider 1's maximization problem is

$$\begin{aligned} \max R &= p_1 (1 - 4p_1) \\ s.t. \frac{1}{4} &\geq p_1 \geq \frac{1}{2} - p_2. \end{aligned}$$

Again, R is concave in p_1 and reaches its maximum at $p_1 = \frac{1}{8}$. Since $\frac{1}{4} \geq p_1 \geq \frac{1}{2} - p_2$, the optimal price in this region is $p_1 = \frac{1}{2} - p_2$ and the revenues are $\left(\frac{1}{2} - p_2\right)(4p_2 - 1)$. For $p_1 > \frac{1}{4}$ we are in profile G where $n_1^p = 0$. Therefore, to find the best response for the case of $1/3 > p_2 \geq 1/4$ we need to compare $\left(2p_2 - \frac{1}{2}\right)(1-2p_2)$ with $\left(\frac{1}{2} - p_2\right)(1-2p_2)$ and $\left(\frac{1}{2} - p_2\right)(4p_2 - 1)$. Among these candidates, $\left(\frac{1}{2} - p_2\right)(1-2p_2)$ generates the highest revenues. Finally, we need to compare these revenues with the revenues from the best equilibrium in profile A, which are smaller than $\left(\frac{1}{2} - p_2\right)(1-2p_2)$, which is the revenue in profile B, and hence the best response for $1/3 > p_2 \geq 1/4$ is $p_1 = \frac{1}{2} - p_2$.

Case 4. $1/2 > p_2 \geq 1/3$. If $p_1 \leq \frac{1}{6}$ we are in profile H where $n_1^p = \frac{1}{2} - p_1$. Provider 1's maximization problem is

$$\begin{aligned} \max R &= p_1\left(\frac{1}{2} - p_1\right) \\ s.t. p_1 &\leq \frac{1}{6}. \end{aligned}$$

Again, R is concave in p_1 and reaches its maximum at $p_1 = \frac{1}{4}$. Since $\frac{1}{4} > \frac{1}{6}$, the optimal price in this region is $p_1 = \frac{1}{6}$ and the revenue is $\frac{1}{18}$. If $\frac{1}{4} \geq p_1 > \frac{1}{6}$ we are in profile E where $n_1^p = 1 - 4p_1$. In profile E provider 1's maximization problem is

$$\begin{aligned} \max R &= p_1(1-4p_1) \\ s.t. \frac{1}{4} &\geq p_1 > \frac{1}{6}. \end{aligned}$$

Again, R is concave in p_1 and reaches its maximum at price $p_1 = \frac{1}{8}$. Since $\frac{1}{8} < \frac{1}{6}$, the optimal price in this region is $p_1 = \frac{1}{6}$ and the revenue is $\frac{1}{18}$. For $p_1 > \frac{1}{4}$ we are in profile G where $n_1^p = 0$. Therefore, the best response for $1/3 > p_2 \geq 1/4$ is $p_1 = \frac{1}{6}$.

We can now summarize the best response of provider 1 as

$$p_1 = \begin{cases} \frac{1}{3}p_2 + \frac{1}{6} & \text{if} \quad p_2 < 1/4 \\ \frac{1}{2} - p_2 & \text{if} \quad 1/3 > p_2 \geq 1/4 \\ \frac{1}{6} & \text{if} \quad 1/2 > p_2 \geq 1/3 \end{cases}.$$

Similarly, the best response of provider 2 is

$$p_2 = \begin{cases} \frac{1}{3}p_1 + \frac{1}{6} & \text{if} \quad p_1 < 1/4 \\ \frac{1}{2} - p_1 & \text{if} \quad 1/3 > p_1 \geq 1/4 \\ \frac{1}{6} & \text{if} \quad 1/2 > p_1 \geq 1/3 \end{cases}.$$

We plot these best responses in Figure 3.

Therefore, in the unique equilibrium of this game both providers announce prices $p_1 = p_2 = \frac{1}{4}$. ■

Duopoly with heterogeneous customers.

Proof of Lemma 4. Assume, by way of contradiction, that there are two different pairs (c_1^*, c_2^*) and $(c_1^{*\prime}, c_2^{*\prime})$ that both satisfy the indifference conditions for the same pair of prices p_1 and p_2. If $c_1^* = c_1^{*\prime}$ (that is, if in both equilibria the same type is indifferent between the priority services of both providers), then $c_2^* = c_2^{*\prime}$, as otherwise the utility of the cutoff type c_1^* from choosing the priority service of provider 2 will not be the

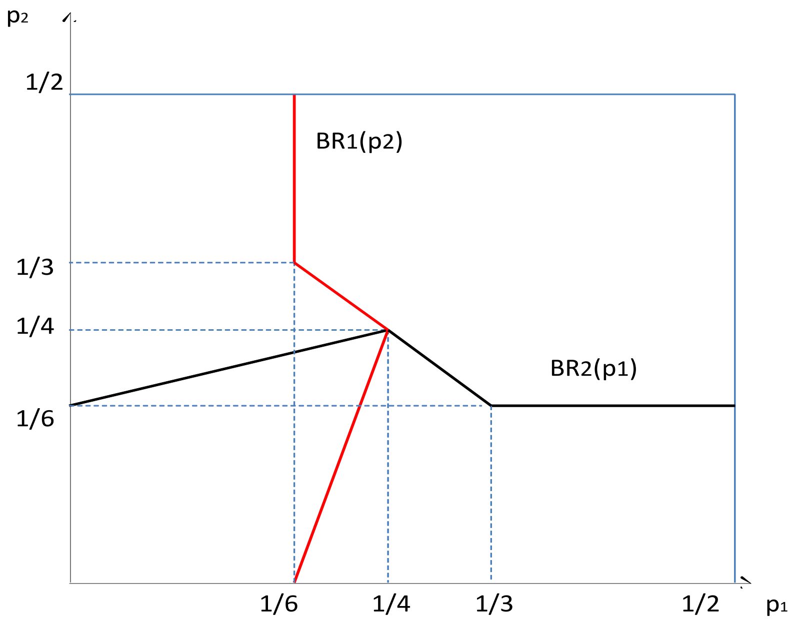


Figure 1: Figure 3. Best responses.

same in the two equilibria, and since the utility from choosing the priority service of provider 1 is the same, this type will not be indifferent between the two priority services of the two providers. This contradicts our assumption that $c_1^* = c_1^{*\prime}$.

Assume now that $c_1^* < c_1^{*\prime}$ (the case of $c_1^* > c_1^{*\prime}$ is similar). The last inequality implies that the utility from joining the priority service of provider 1 is higher (for all types) in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$ than in equilibrium (c_1^*, c_2^*) since the prices are the same, but in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$ fewer customers join the priority service of provider 1 than in the equilibrium given by (c_1^*, c_2^*) (as $F(c_1^*) < F(c_1^{*\prime})$). As $c_1^* < c_1^{*\prime}$, it follows that, in equilibrium (c_1^*, c_2^*), type $c_1^{*\prime}$ prefers the priority service of provider 1 to the priority service of provider 2. Since in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$ type $c_1^{*\prime}$ is indifferent between the two priority services of the two providers, it must be the case that the utility of joining the priority service of provider 2 is higher in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$ than in the equilibrium given by (c_1^*, c_2^*). Therefore,

$$F(c_1^{*\prime}) - F(c_2^{*\prime}) < F(c_1^*) - F(c_2^*).$$

This inequality, together with $F(c_1^*) < F(c_1^{*\prime})$, implies that $F(c_2^{*\prime}) > F(c_2^*)$. It further implies that $c_2^{*\prime} > c_2^*$. However, it also implies that the utility from non-priority services is lower in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$ than in equilibrium (c_1^*, c_2^*), since in $(c_1^{*\prime}, c_2^{*\prime})$ fewer agents join the priority services of both providers. This in turn implies that the type that is indifferent between the priority service of provider 2 and the regular service of any provider must be lower in equilibrium (c_1^*, c_2^*) than in equilibrium $(c_1^{*\prime}, c_2^{*\prime})$. This contradicts our assumption that $c_2^{*\prime} > c_2^*$. ■

Proof of Proposition 13. Assume that one of the providers, say provider 2, sets

$p_2 = 0$. We will show that provider 1 should not set its price for priority service to 0. More precisely, provider 1 can increase its profits by setting $p_1 > 0$. First observe that from (7) it follows that $c_2^* = 0$. Plugging it into (6) gives us

$$p_1 = c_1^* \left[F\left(c_1^*\right) - \frac{1}{2} \right].$$

Hence, setting $p_1 = 0$ implies that $c_1^* = F^{-1}\left(\frac{1}{2}\right)$, that is, provider 1 serves half of the market. While setting a price $p_1 \in \left(0, \frac{\bar{c}}{2}\right)$ the provider will serve a positive measure of the market at a positive price, and hence gets a positive profit.

Now we show that $p_2 = 0$ and $p_1 > 0$ is not an equilibrium. We show that provider 2 can increase its profit by setting a positive price. If $p_1 < \frac{\bar{c}}{4}$, then setting $p_2 = p_1$ implies that all customers with waiting costs above $4p_1$ will acquire priority service from one of the providers and provider 2 will get a profit of $p_1 \frac{1-F(4p_1)}{2} > 0$. If $p_1 \geq \frac{\bar{c}}{4}$, then, since $0 \leq \frac{1-F\left(c_1^*\right)}{2} \leq \frac{1}{2}$ and both c_1^* and c_2^* are continuous in p_2 and p_1, there exists $p_2 > 0$ close enough to zero (and so $p_2 < p_1$) such that the resulting cutoff type c_2^* satisfies $0 < c_2^* < c_1^*$, and hence provider 2 gets a positive profit. ■

Proof of Proposition 14. We show that for any $p_1 \geq p_2 > 0$ holds

$$-\frac{\mathbb{E}(c)}{4} \geq -\int_{c_1^*}^{\bar{c}} \left(p_1 + c\frac{1 - F\left(c_1^*\right)}{2} \right) f(c)dc \tag{15}$$

$$-\int_{c_2^*}^{c_1^*} \left(p_2 + c\frac{F\left(c_1^*\right) - F\left(c_2^*\right)}{2} \right) f(c)dc - \int_0^{c_2^*} c\left(1 - F\left(c_1^*\right) + \frac{n_1^{np}}{2} \right) f(c)dc$$

Observe that since $p_1 \geq p_2$ it implies that

$$F\left(c_1^*\right) - F\left(c_2^*\right) \geq 1 - F\left(c_1^*\right).$$

Furthermore, $F\left(c_1^*\right) > \frac{1}{2}$.

Plugging the expressions for p_1, p_2 and n_1^{np} we can rewrite the right hand side of the last inequality as

$$-\frac{1 - F\left(c_1^*\right)}{2}\mathbb{E}(c) - \frac{F\left(c_1^*\right) - F\left(c_2^*\right) - 1 + F\left(c_1^*\right)}{2} \int_0^{c_1^*} cf(c)dc$$

$$-\left(\frac{1 - F\left(c_1^*\right)}{2} + \frac{F\left(c_2^*\right)}{4} \right) \int_0^{c_2^*} cf(c)dc - \frac{c_2^* - c_1^*}{2}\left(1 - F\left(c_1^*\right)\right)$$

$$-\left(c_1^* - \frac{c_2^*}{2} \right) \left(F\left(c_1^*\right) - \frac{F\left(c_2^*\right)}{2} \right) \left(1 - F\left(c_1^*\right)\right) - c_2^*\left(F\left(c_1^*\right) - F\left(c_2^*\right)\right) \left(\frac{1 - F\left(c_1^*\right)}{2} + \frac{F\left(c_2^*\right)}{4} \right)$$

The derivative of the last expression with respect to c_1^* is

$$f\left(c_1^*\right)\left(\frac{\mathbb{E}(c)}{2} - \int_0^{c_1^*} cf(c)dc + \frac{1}{2}\int_0^{c_2^*} cf(c)dc - c_1^*\left(1 - F\left(c_1^*\right)\right) + \frac{c_2^*}{2}\left(1 - F\left(c_2^*\right)\right)\right.$$

$$\left. -\frac{1 - F\left(c_1^*\right)}{f\left(c_1^*\right)} \left[F\left(c_1^*\right) - \frac{1}{2} - \frac{F\left(c_2^*\right)}{2} \right]\right)$$

We will first show that for the relevant parameters this derivative is negative. Plugging the expressions for $F(c)=c^{\theta}$ and $f(c)=\theta c^{\theta-1}$ gives

$$\begin{aligned}
& \frac{\theta}{\theta+1}\frac{1}{2}-\frac{\theta}{\theta+1}\left(c_1^*\right)^{\theta+1}-c_1^*\left(1-\left(c_1^*\right)^{\theta}\right)+\frac{1}{2}\frac{\theta}{\theta+1}\left(c_2^*\right)^{\theta+1} \\
& +\frac{1}{2}c_2^*\left(1-\left(c_2^*\right)^{\theta}\right)-\frac{1-\left(c_1^*\right)^{\theta}}{\theta\left(c_1^*\right)^{\theta-1}}\left(\left(c_1^*\right)^{\theta}-\frac{1}{2}-\frac{\left(c_2^*\right)^{\theta}}{2}\right) \\
= & \left(1-\frac{1}{\theta+1}\right)\left(\frac{1}{2}-\left(c_1^*\right)^{\theta+1}+\frac{1}{2}\left(c_2^*\right)^{\theta+1}\right)-c_1^*\left(1-\left(c_1^*\right)^{\theta}\right) \\
& +\frac{1}{2}c_2^*\left(1-\left(c_2^*\right)^{\theta}\right)+\frac{1-\left(c_1^*\right)^{\theta}}{\theta\left(c_1^*\right)^{\theta-1}}\left(\frac{1}{2}-\left(c_1^*\right)^{\theta}+\frac{\left(c_2^*\right)^{\theta}}{2}\right) \\
= & \frac{1}{2}-c_1^*+\frac{1}{2}c_2^*-\frac{1}{\theta+1}\left(\frac{1}{2}-\left(c_1^*\right)^{\theta+1}+\frac{1}{2}\left(c_2^*\right)^{\theta+1}\right) \\
& +\frac{1-\left(c_1^*\right)^{\theta}}{\theta\left(c_1^*\right)^{\theta-1}}\left(\frac{1}{2}-\left(c_1^*\right)^{\theta}+\frac{\left(c_2^*\right)^{\theta}}{2}\right)
\end{aligned}$$

This expression increases in c_2^*. Hence, to show that the last expression is negative, it is enough to show it for c_2^* s.t.

$$\left(c_2^*\right)^{\theta}=2\left(c_1^*\right)^{\theta}-1.$$

Plugging the expression for which c_2^* we get that it is enough to show that

$$\frac{1}{2}-c_1^*+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{1}{\theta}}-\frac{1}{\theta+1}\left(\frac{1}{2}-\left(c_1^*\right)^{\theta+1}+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{\theta+1}{\theta}}\right)<0 \text{ for any } \left(c_1^*\right)^{\theta}>\frac{1}{2} \text{ and } \theta\geq 1. \tag{16}$$

First observe that

$$\begin{aligned}
& \frac{1}{2}-c_1^*+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{1}{\theta}} \\
= & -\left(c_1^*-\frac{1}{2}\right)+\left(\frac{1}{2}\right)^{\frac{\theta-1}{\theta}}\left(\left(c_1^*\right)^{\theta}-\frac{1}{2}\right)<0
\end{aligned}$$

Where the last inequality holds since $\theta\geq 1$ and $1>\left(c_1^*\right)^{\theta}\geq\frac{1}{2}$. If $\frac{1}{2}-\left(c_1^*\right)^{\theta+1}+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{\theta+1}{\theta}}>0$ for any $\left(c_1^*\right)^{\theta}>\frac{1}{2}$ and $\theta\geq 1$ we have inequality (16). Otherwise, since $\theta\geq 1$ it is enough to show that

$$\begin{aligned}
\frac{1}{2}-c_1^*+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{1}{\theta}}-\frac{1}{2}\left(\frac{1}{2}-\left(c_1^*\right)^{\theta+1}+\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{\theta+1}{\theta}}\right) &< 0 \\
2\left(\frac{1}{2}-c_1^*\right)+2\left(c_1^*\right)^{\theta}-1-\frac{1}{2}+\left(c_1^*\right)^{\theta+1}-\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{\theta+1}{\theta}} &< 0 \\
-2c_1^*+2\left(c_1^*\right)^{\theta}-\frac{1}{2}+\left(c_1^*\right)^{\theta+1}-\frac{1}{2}\left(2\left(c_1^*\right)^{\theta}-1\right)^{\frac{\theta+1}{\theta}} &< 0
\end{aligned}$$

where the last inequality holds since its left hand side is strictly increasing in c_1^* and for $c_1^* = 1$ its left hand side equals to 0.

Hence, to show (15) it is enough to show it for $(c_1^*)^\theta = \frac{1}{2}$ and $c_2^* = \left(2(c_1^*)^\theta - 1\right)^{\frac{1}{\theta}} = 0$. However, plugging these expressions into (15) gives that it holds as equality. ■

Duopoly with heterogeneous customers. Derivations of example.

We analyze here the duopoly equilibrium for heterogeneous customers. Given that $c_1^*(p_1, p_2)$ and $c_2^*(p_1, p_2)$ are solutions to (6) and (7), the profit of provider 1 if it sets a price of p_1 for priority service and provider 2 sets price of p_2 is

$$\pi_1(p_1, p_2) = p_1(1 - F(c_1^*(p_1, p_2)))$$

and, similarly, the profit of provider 2 is

$$\pi_2(p_1, p_2) = p_2(F(c_1^*(p_1, p_2)) - F(c_2^*(p_1, p_2))).$$

For a profile (p_1, p_2) to be an equilibrium, it must be the case that[34]

$$\frac{\partial \pi_1(p_1, p_2)}{\partial p_1} = 0 \text{ and } \frac{\partial \pi_2(p_1, p_2)}{\partial p_2} = 0.$$

Hence the first-order conditions are

$$\begin{aligned}(1 - F(c_1^*(p_1, p_2))) - p_1 f(c_1^*(p_1, p_2)) \frac{\partial c_1^*(p_1, p_2)}{\partial p_1} &= 0 \\ (F(c_1^*(p_1, p_2)) - F(c_2^*(p_1, p_2))) + p_2\left(f(c_1^*(p_1, p_2)) \frac{\partial c_1^*(p_1, p_2)}{\partial p_2} - f(c_2^*(p_1, p_2)) \frac{\partial c_2^*(p_1, p_2)}{\partial p_2}\right) &= 0.\end{aligned}$$

We now calculate $\frac{\partial c_1^*(p_1,p_2)}{\partial p_1}$, $\frac{\partial c_1^*(p_1,p_2)}{\partial p_2}$ and $\frac{\partial c_2^*(p_1,p_2)}{\partial p_2}$ using the implicit function theorem. Denote by G_1 and G_2 as follows

$$\begin{aligned}G_1(p_1, p_2, c_1, c_2) &= p_1 - p_2 - c_1 \frac{2F(c_1) - F(c_2) - 1}{2} \\ G_2(p_1, p_2, c_1, c_2) &= p_2 - c_2 \frac{2 - 2F(c_1) + F(c_2)}{4}.\end{aligned}$$

The implicit function theorem implies that the derivatives $\frac{\partial c_j^*(p_1,p_2)}{\partial p_i}$ can be calculated from

$$\begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix} \times \begin{pmatrix} \frac{\partial c_1}{\partial p_i} \\ \frac{\partial c_2}{\partial p_i} \end{pmatrix} = -\begin{pmatrix} \frac{\partial G_1}{\partial p_i} \\ \frac{\partial G_2}{\partial p_i} \end{pmatrix},$$

conditional that

$$\det \begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix} \neq 0,$$

[34] In the case of a symmetric equilibrium the FOC are $\frac{\partial \pi_1(p,p)}{\partial p_1} \leq 0$ and $\frac{\partial \pi_2(p,p)}{\partial p_2} \geq 0$.

where the derivatives are evaluated at $(p_1, p_2, c_1^*(p_1, p_2), c_2^*(p_1, p_2))$. Applying Cramer's rule we get

$$\frac{\partial c_1}{\partial p_1} = -\frac{\det\begin{pmatrix} \frac{\partial G_1}{\partial p_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial p_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix}}{\det\begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix}}, \frac{\partial c_1}{\partial p_2} = -\frac{\det\begin{pmatrix} \frac{\partial G_1}{\partial p_2} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial p_2} & \frac{\partial G_2}{\partial c_2} \end{pmatrix}}{\det\begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix}}, \frac{\partial c_2}{\partial p_2} = -\frac{\det\begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial p_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial p_2} \end{pmatrix}}{\det\begin{pmatrix} \frac{\partial G_1}{\partial c_1} & \frac{\partial G_1}{\partial c_2} \\ \frac{\partial G_2}{\partial c_1} & \frac{\partial G_2}{\partial c_2} \end{pmatrix}}$$

where

$$\begin{aligned} \frac{\partial G_1}{\partial c_1} &= -\frac{2F(c_1) - F(c_2) - 1}{2} - c_1 f(c_1), \frac{\partial G_1}{\partial c_2} = c_1 \frac{f(c_2)}{2} \\ \frac{\partial G_2}{\partial c_1} &= c_2 \frac{f(c_1)}{2}, \frac{\partial G_2}{\partial c_2} = -\frac{2 - 2F(c_1) + F(c_2)}{4} - c_2 \frac{f(c_2)}{4} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial G_1}{\partial p_1} &= 1, \frac{\partial G_1}{\partial p_2} = -1 \\ \frac{\partial G_2}{\partial p_1} &= 0, \frac{\partial G_2}{\partial p_2} = 1. \end{aligned}$$

Assuming distribution function $F(c) = c^\theta$ gives the following first-order conditions

$$\begin{aligned} \left(1 - (c_1)^\theta\right) - p_1 \theta (c_1)^{\theta-1} \frac{\partial c_1^*(p_1, p_2)}{\partial p_1} &= 0 \\ \left((c_1)^\theta - (c_2)^\theta\right) + p_2 \left(\theta (c_1)^{\theta-1} \frac{\partial c_1^*(p_1, p_2)}{\partial p_2} - \theta (c_2)^{\theta-1} \frac{\partial c_2^*(p_1, p_2)}{\partial p_2}\right) &= 0. \end{aligned}$$

with

$$\begin{aligned} p_1 &= c_2 \left[\frac{1 - (c_1)^\theta}{2} + \frac{(c_2)^\theta}{4}\right] + c_1 \frac{2(c_1)^\theta - (c_2)^\theta - 1}{2} \\ p_2 &= c_2 \left[\frac{1 - (c_1)^\theta}{2} + \frac{(c_2)^\theta}{4}\right] \end{aligned}$$

and

$$\begin{aligned} \frac{\partial c_1}{\partial p_1} &= \frac{2 - 2(c_1)^\theta + (1+\theta)(c_2)^\theta}{\frac{2(1+\theta)(c_1)^\theta - (c_2)^\theta - 1}{2}\left(2 - 2(c_1)^\theta + (1+\theta)(c_2)^\theta\right) - \theta^2 (c_1 c_2)^\theta} \\ \frac{\partial c_1}{\partial p_2} &= \frac{2\theta c_1 (c_2)^{\theta-1} - \left(2 - 2(c_1)^\theta + (1+\theta)(c_2)^\theta\right)}{\frac{2(1+\theta)(c_1)^\theta - (c_2)^\theta - 1}{2}\left(2 - 2(c_1)^\theta + (1+\theta)(c_2)^\theta\right) - \theta^2 (c_1 c_2)^\theta} \\ \frac{\partial c_2}{\partial p_2} &= \frac{4(1+\theta)(c_1)^\theta - 2(c_2)^\theta - 2 - 2\theta c_2 (c_1)^{\theta-1}}{\frac{2(1+\theta)(c_1)^\theta - (c_2)^\theta - 1}{2}\left(2 - 2(c_1)^\theta + (1+\theta)(c_2)^\theta\right) - \theta^2 (c_1 c_2)^\theta} \end{aligned}$$

Example 3. Plugging $\theta = 1/2$ into the first order condition gives us

$$c_1 = 0.67336,\ c_2 = 0.34744.$$

This implies that the prices are

$$\begin{aligned} p_1 &= c_2 \left(\frac{1-\sqrt{c_1}}{2} + \frac{\sqrt{c_2}}{4} \right) + c_1 \frac{2\sqrt{c_1} - \sqrt{c_2} - 1}{2} = 0.0.09978 \\ p_2 &= c_2 \left(\frac{1-\sqrt{c_1}}{2} + \frac{\sqrt{c_2}}{4} \right) = 0.08237 \end{aligned}$$

and the providers' profits are

$$\begin{aligned} \pi_1 &= p_1 \left(1 - \sqrt{c_1}\right) = 0.0179 \\ \pi_2 &= p_2 \left(\sqrt{c_1} - \sqrt{c_2}\right) = 0.019. \end{aligned}$$

The expected waiting time for regular service is

$$1 - F(c_1) + \frac{n_1^{np}}{2} = 0.35264.$$

Hence, the customers' surplus is

$$-0.35264 \int_0^{c_2} c dc + \int_{c_2}^{c_1} \left(-p_2 - \frac{n_2^p}{2} c \right) dc + \int_{c_1}^{1} \left(-p_1 - \frac{n_1^p}{2} c \right) dc = -0.0878.$$

Without priority the customers' surplus is

$$-\frac{\mathbb{E}c}{4} = -\frac{\int_0^1 \frac{1}{2}\sqrt{s} ds}{4} = -0.083.$$

References

[1] Acemoglu, D., and A. Ozdaglar (2007), "Competition and Efficiency in Congested Markets," *Mathematics of Operations Research,* **32**, 1-31.

[2] Aghion, P., Dewatripont, M., and P. Rey (1999), "Competition, Financial Discipline, and Growth," *Review of Economic Studies*, **66**, 825-852.

[3] Aghion, P., Harris, C., Howitt, P.,and J. Vickers (2001), "Competition, Imitation and Growth with Step-by-Step Innovation," *Review of Economic Studies*, **68**, 467-492.

[4] Aghion, P., Harris, C., and J. Vickers (1997), "Competition and Growth with Step-by-Step Innovation: An Example," *European Economic Review, Paper and Proceedings,* **41**, 771–782.

[5] Ashenfelter, O and D. Genesove (1992) "Testing Price Anomalies in Real-Estate Auctions" *The American Economic Review Papers and Proceedings* **82**, 501-505.

[6] Bagnoli, M., and T. Bergstrom (2005), "Log-concave Probability and Its Applications," *Economic Theory* **26**, 445-469.

[7] Bernheim, D., and M. Whinston (1998), "Exclusive Dealing," *Journal of Political Economy*, **106**, 64-103.

[8] Budish, E., Cramton, P., and J. Shim (2015), "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *Quarterly Journal of Economics,* **130**, 1547–1621.

[9] Bulow, J., and P. Klemperer (2012), "Regulated Prices, Rent Seeking, and Consumer Surplus," *Journal of Political Economy* **120**, 160-186.

[10] Bulow, J., and J. Roberts (1989), "The Simple Economics of Optimal Auctions," *Journal of Political Economy,* **97**, 1060-1090.

[11] Chao H., and Wilson R. (1987), "Priority Service: Pricing, Investment and Market Organization," *American Economic Review,* **77**, 899-916.

[12] Chun, Y., Mitra, M., and S. Mutuswami (2019), "Recent Developments in the Queueing Problem," *TOP* (Spanish Society of Statistics and Operations Research), 27, 1-23.

[13] Compte, O., Jenny, F., and P. Rey (2002), "Capacity Constraints, Mergers and Collusion," *European Economic Review*, **46**, 1-29.

[14] Cramton, P., Gibbons, R., and P. Klemperer (1987), "Dissolving a Partnership Efficiently," *Econometrica*, **55**, 615-632.

[15] De Borger, B., and K. Van Dender (2006), "Prices, Capacities and Service Levels in a Congestable Bertrand Duopoly," *Journal of Urban Economics*, **60**, 264-283.

[16] Deneckere, R., and P. McAfee (1996), "Damaged Goods," *Journal of Economics and Management Strategy,* **5**, 149-174.

[17] Dolan, R. (1978), "Incentives Mechanisms for Priority Queuing Problems," *The Bell Journal of Economics,* **9**, 421-436.

[18] Dworczak, P. Kominers, S. D., and M. Akbarpour (2021), "Redistribution Through Markets," *Econometrica*, **89**, 1665-1698.

[19] Edelson, N (1971), "Congestion Tolls under Monopoly," *American Economic Review,* **61**, 873-882.

[20] Einav, L., Kuchler, T., Levin, J., and N. Sundaresan (2015) "Assessing Sale Strategies in Online Markets Using Matched Listings," *American Economic Journal: Microeconomics* **7,** 215–247.

[21] Gershkov, A. and P. Schweinzer (2010), "When Queueing is Better than Push and Shove," *International Journal of Game Theory*, **39**, 409-430.

[22] Glazer, A. and R. Hassin (1986), "Stable Priority Purchasing in Queues," *Operations Research Letters* **4**, 285-288.

[23] Gomes, R., and J. Tirole (2018), "Missed Sales and the Pricing of Ancillary Goods," *The Quarterly Journal of Economics*, **133**, 2097-2169.

[24] Hall, J (2018) "Pareto Improvements from Lexus Lanes: The Effects of Pricing a Portion of the Lanes on Congested Highways," *Journal of Public Economics*, **158**, 113-125.

[25] Harrington, J. (2017), *The Theory of Collusion and Competition Policy*, The MIT Press.

[26] Hassin R., and Haviv, M (2003), *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, Springer.

[27] Haviv, M. and E. Winter (2020), "An Optimal Mechanism Charging for Priority in a Queue," *working paper.*

[28] Hoppe, H., Moldovanu, B., and E. Ozdenoren (2011), "Coarse Matching with Incomplete Information," *Economic Theory*, 47, 73-104.

[29] Kittsteiner, T., and B. Moldovanu (2005), "Priority Auctions and Queue Disciplines that Depend on Processing Time," *Management Science*, **51**, 236-248.

[30] Levhari, D. and I. Luski (1978), "Duopoly Pricing and Waiting Lines," *European Economic Review,* **11**, 17-35.

[31] Luski, I (1976), "On Partial Equilibrium in a Queuing System with Two Servers," *Review of Economic Studies*, **43**, 519-525.

[32] McAfee, P. (2002), "Coarse Matching," *Econometrica,* **70**, 2025-2034.

[33] Mussa, M. and S. Rosen (1978), "Monopoly and Product Quality," *Journal of Economic Theory*, **18**, 301-317.

[34] Reitman, D (1991), "Endogenous Quality Differentiation in Congested Markets," *Journal of Industrial Economics*, **39**, 621-647.

[35] Segal, I. (2003), "Coordination and Discrimination in Contracting with Externalities: Divide and Conquer?," *Journal of Economic Theory*, **113**, 147-181.

[36] Wilson, R. (1989), "Efficient and Competitive Rationing," *Econometrica,* **57**, 1-40.

[37] Winter, E. (2004), "Incentives and Discrimination," *American Economic Review*, **94**, 764-773.