A Mechanism Design Approach to "Gainers and Losers in Priority Services"

Peiran Xiao*

April 2024

Gershkov and Winter (2023, henceforth GW) study a model of priority service (PS) and analyze its welfare implications on consumers. They show that introducing PS can be detrimental to consumer welfare when the monopoly service provider extracts more revenue than the efficiency gains. For multiple priority levels, they find that if the distribution has an increasing failure rate (IFR), the provider's revenue is strictly increasing in the number of priority levels, whereas the customers' welfare is higher if PS is not offered than if multiple priority levels are offered.

In this note, I reformulate their model (with a monopoly service provider) as a mechanism design problem under a feasibility condition. Thus, I rewrite the revenue and consumer welfare maximization problems as linear optimization problems under majorization constraints. Using the techniques in Kleiner, Moldovanu, and Strack (2021), I provide the necessary and sufficient conditions for GW's Propositions 1, 2, 7, and 8, while allowing for stochastic priority levels. In particular, consumer welfare is decreasing (increasing) in the number of priority levels if and only if the failure rate is increasing (decreasing), whereas an increasing failure rate is sufficient (but not necessary) for the provider's revenue to be increasing in the number of priority levels. Thus, the trade-off between the provider's revenue and consumer welfare is less stark—for some distributions with decreasing failure rates (e.g., exponential and Weibull distributions), increasing the number of priority levels can increase both. Infinitely many priority levels can be implemented by an all-pay auction. This approach can also be applied to the model of contests for status by Moldovanu, Sela, and Shi (2007).

*Department of Economics, Boston University. Email: pxiao@bu.edu.

1 Model

In their benchmark model, a monopoly service provider faces a continuum of customers of mass 1 with heterogeneous unit costs of waiting (types), denoted by c, which has a distribution F with bounded support $[0, \bar{c}]$ and density $f(c) > 0$. A customer with type c who gets service at time t and pays $p \geq 0$ has a utility of $-p - tc$. The provider can serve a single customer at each instant. They normalize the service time of a mass m of consumers to be exactly m units of time and the cost of the provider to be zero, which implies $t \in [0, 1]$ because the total mass of customers is 1.

They assume the price of the regular service is zero so that there is no exclusion in the benchmark model. Then, in Section IV, they allow for endogenous pricing of the regular service (and thus exclusion) by introducing the value of the service to the consumer. I maintain this assumption first and consider endogenous pricing of the regular service and exclusion in Section 4, as in GW's Section IV.

2 A Mechanism Design Approach

It is without loss to consider a direct mechanism $\{p(c), t(c)\}$. One can view the direct mechanism as consisting of an ex-post payment $\mathbf{p}_i \colon [0, \bar{c}]^n \to \mathbb{R}_+$ and a (potentially random) ex-post waiting time $\mathbf{t} \colon [0, \bar{c}]^n \times \Omega \to [0, 1]^n$, where n is the number of consumers; and $p(c) \equiv p_i(c_i) = \mathbf{E}_{c_{-i}}[\mathbf{p}_i(c_i, c_{-i})|c_i]$ and $t(c) \equiv t_i(c_i) = \mathbf{E}_{c_{-i}}[\mathbf{t}_i(c_i, c_{-i}, \omega)|c_i]$ are the interim expected payment and waiting time. Denote $U(c) = -p(c) - t(c)c$.

Lemma 1. *A direct mechanism $\{p(c), t(c)\}$ is incentive-compatible if and only if*

- $U(c) = U(0) - \int_0^c t(x)\,\mathrm{d}x$ *for all* $c \in [0, \bar{c}]$*, and*
- $t(c)$ *is decreasing.*

Alternative interpretation. The provider designs a queuing scheme $\ell \colon [0, \bar{c}] \to \Delta L$ that maps the consumer's type c to a (possibly stochastic) priority level $\ell(c)$ along with a payment $p \colon [0, \bar{c}] \to \mathbb{R}_+$; the higher the priority level, the less time he needs to wait.[1] If there are $k \geq 1$ priority levels, $L = \{1, \ldots, k\}$. If there are infinitely many priority levels, $L = [0, 1]$. When priority levels are *deterministic*, $\ell(c) \in L$ is degenerate. By the assumptions on the service time, for a given level $\ell_0 \in L$, the expected waiting time (conditional on ℓ_0) is $\tau(\ell_0) = \Pr(\ell(c) > \ell_0) + \Pr(\ell(c) = \ell_0)/2$. For deterministic priority levels, incentive compatibility requires that $\ell(c)$ must be (weakly) increasing, so the

[1]This can be implemented by a pricing scheme as a function of the priority level.

consumer's expected waiting time is $t(c) = 1 - \tau(\ell(c)) = 1 - \mathbf{E}[F(\tilde{c})|\ell(\tilde{c}) = \ell(c)]$. For potentially stochastic priority levels, the expected waiting time is $t(c) = 1 - \mathbf{E}_{\tilde{\ell}\sim\ell(c)}[\tau(\tilde{\ell})] = 1 - \mathbf{E}_{\tilde{\ell}\sim\ell(c)}[\mathbf{E}[F(\tilde{c})|\tilde{\ell}]]$.[2]

A decreasing waiting time function $t(c)$, albeit incentive-compatible, is not necessarily feasible because it may be unable to be induced by an ex-post allocation (or queuing scheme). Formally, say $t(c)$ is *feasible* if there exists an ex-post allocation $\mathfrak{t}$ (or a queuing scheme ℓ) that induces $t(c)$ as the interim waiting time. Now I characterize the necessary and sufficient condition for a decreasing $t(c)$ to be feasible.

It is convenient to denote the consumer's priority value by $s(c) = 1 - t(c)$, which quantifies priority by the time he saved compared to being served last. In other words, $s(c) = \mathbf{E}_{\tilde{\ell}\sim\ell(c)}[\mathbf{E}[F(\tilde{c})|\tilde{\ell}]]$, where $\ell(c)$ is his (relative) priority *level*.

Theorem 2 (Feasibilty)**.** *A decreasing* $t(c)$ *is feasible if and only if* $s(c) = 1 - t(c)$ *is a mean-preserving spread of* $F(c)$ *in the quantile space, denoted by* $s \in \mathrm{MPS}(F)$*, that is,*

$$\int_0^x s(c)\,\mathrm{d}F(c) \geq \int_0^x F(c)\,\mathrm{d}F(c) \text{ for all } x \in [0, \bar{c}], \tag{1}$$

$$\int_0^{\bar{c}} s(c)\,\mathrm{d}F(c) = \int_0^{\bar{c}} F(c)\,\mathrm{d}F(c) = 1/2. \tag{2}$$

Proof sketch. À la Kleiner, Moldovanu, and Strack (2021, Theorem 3). □

Remark 1. The condition resembles the symmetric version of Border's Theorem in reduced-form auctions (see Maskin and Riley, 1984; Matthews, 1984; Border, 1991).

Remark 2. If the priority levels $\ell(c)$ are deterministic, $s(c) = 1 - t(c)$ can only be an extreme point of $\mathrm{MPS}(F)$ (either the majorization constraint or the monotonicity constraint binds) (see Kleiner, Moldovanu, and Strack, 2021, Theorem 1). Since I allow for stochastic priority levels, $t(c)$ can take other forms.

Remark 3. An analog of the theorem also applies to Moldovanu, Sela, and Shi (2007) with a continuum of agents, in which the feasibility condition on the value of status can be written as $s(\theta) \in \mathrm{MPS}(2F - 1)$, and the total value is normalized to $\mathbf{E}[s] = \mathbf{E}[2F - 1] = 0$.

The intuition is that the provider can induce full separation (i.e., $t(c) = 1 - F(c)$) by offering infinitely many priority levels and serving every type in the descending order of their costs. Any pooling of different types into the same priority level makes $s \equiv 1 - t$ a mean-preserving spread of F.

[2] For deterministic priority levels, it can be viewed as a monotone categorization problem (see Rayo (2013) and Onuchic and Ray (2023)). See also Xiao (2024) for potentially stochastic levels.

Example 2.1. Suppose there are two priority levels—regular service ($\{p_L, t_L\}$) and priority service ($\{p_H, t_H\}$), where $p_L = 0$ (no exclusion). Denote the cutoff type by c^*. Then, the waiting time is $t_L = 1 - F(c^*)/2$ for $c < c^*$ and $t_H = (1 - F(c^*))/2$ for $c \geq c^*$. It is easy to check that $\mathbf{E}[t] = 1/2$ and that $\int_0^c t(c')\,\mathrm{d}F(c') \leq \int_0^c (1 - F(c'))\,\mathrm{d}F(c')$ for all $c \in [0, \bar{c}]$ (with equality at c^* and $\bar{c}$).

3 Optimal Mechanism

Two extreme mechanisms are particularly of interest: full separation ($t(c) = 1 - F(c)$) and total pooling ($t(c) = 1/2$)[3] ; the former can be induced by offering infinitely many priority service levels, and the latter can be induced by offering no priority service. Note that I allow for stochastic priority levels.

For this section, I maintain GW's assumption in the benchmark model that the regular service $\{p_L, t_L\}$ is free ($p_L = 0$) so that there is no exclusion. Thus, the type-c consumer's reservation utility is $\underline{U}(c) = -t_L c \leq -tc$, where t_L is the expected waiting time of the regular service, which is endogenously determined by the scheme $t(c)$.

3.1 Revenue maximization

I first consider the optimal mechanism that maximizes the provider's revenue. The revenue maximization problem is given by

$$\max_{t(c),p(c)} \int_0^{\bar{c}} p(c)\,\mathrm{d}F(c) \tag{3}$$

subject to

$$p(c) \geq 0 \tag{4}$$

$$U(c) \equiv -p(c) - t(c)c \geq -t_L c \qquad \text{(IR)} \tag{5}$$

$$-p(c) - t(c)c = U(0) - \int_0^c t(x)\,\mathrm{d}x \qquad \text{(IC)} \tag{6}$$

$$t(c) \text{ is decreasing} \tag{7}$$

$$s \equiv 1 - t \in \mathrm{MPS}(F) \qquad \text{(MPS)} \tag{8}$$

[3]Both are the extreme points of MPS(F).

The revenue is given by

$$R = \int_0^{\bar{c}} p(c)\,\mathrm{d}F(c) = \int_0^{\bar{c}} \left(\frac{1-F(c)}{f(c)} - c\right) t(c)\,\mathrm{d}F(c) - U(0). \tag{9}$$

Because $U(0) = -p(0) \geq 0$ (IR), it is optimal to set $U(0) = p(0) = 0$ (even if p were allowed to be negative). Denote $J(c) = c - \frac{1-F(c)}{f(c)}$. The revenue maximization problem is equivalent to

$$\max_{s \in \mathrm{MPS}(F)} \int_0^{\bar{c}} J(c) s(c)\,\mathrm{d}F(c) \tag{10}$$

where $s(c) = 1 - t(c)$ is increasing.

Note that I maintain the assumption that the regular service is free so that there is no exclusion. Kleiner, Moldovanu, and Strack (2021, Proposition 2) immediately implies the following results.

Proposition 1 (Cf. GW's Proposition 8)**.** *The provider's revenue is (strictly) increasing in the number of priority levels if and only if* $c - \frac{1-F(c)}{f(c)}$ *is (strictly) increasing.*

Remark 4. GW's Proposition 8 provides a sufficient condition for this result: F satisfies the IFR property (i.e., $\frac{1-F(c)}{f(c)}$ is decreasing).

Corollary 1.1. *The revenue-maximizing mechanism offers infinitely many priority levels (i.e., full separation) if and only if* $c - \frac{1-F(c)}{f(c)}$ *is increasing, which is payoff equivalent to an all-pay auction.*

Indeed, infinitely many priority levels can be implemented by an all-pay auction, in which the more money a consumer pays, the more ahead he is in the line (and the less time he needs to wait).

3.2 Consumer welfare maximization

The consumer welfare maximization problem is given by

$$\max_{t(c), p(c)} \int_0^{\bar{c}} U(c)\,\mathrm{d}F(c) \tag{11}$$

subject to constraints (4)–(8). Analogous to the previous derivations, the customer's welfare can be written as

$$W = \int_0^{\bar{c}} U(c)\,\mathrm{d}F(c) = \int_0^{\bar{c}} \left(-\frac{1-F(c)}{f(c)}\right) t(c)\,\mathrm{d}F(c) + U(0). \tag{12}$$

Because $p(0) \geq 0$, $U(0) = -p(0) \geq 0$ (IR) implies $U(0) = p(0) = 0$.[4] The revenue maximization problem is equivalent to

$$\max_{s \in \mathrm{MPS}(F)} \int_0^{\bar{c}} \left(\frac{1 - F(c)}{f(c)} \right) s(c) \,\mathrm{d}F(c) - \mathbf{E}[c] \tag{13}$$

where $s(c) = 1 - t(c)$ is increasing. Kleiner, Moldovanu, and Strack (2021, Proposition 2) immediately implies the following results.

Proposition 2 (Cf. GW's Proposition 7)**.** *The customers' welfare is increasing (decreasing) in the number of priority levels if and only if $\frac{1-F(c)}{f(c)}$ is increasing (decreasing). Without priority service, the customers' welfare is $-\mathbf{E}[c]/2$.*

Remark 5. GW's Proposition 7 shows the sufficiency of IFR property ($\frac{1-F(c)}{f(c)}$ is decreasing) for no priority service to be consumer welfare-maximizing.

Corollary 2.1. *The consumer welfare-maximizing mechanism*

(i) offers no priority service (i.e, total pooling) if and only if $\frac{1-F(c)}{f(c)}$ is decreasing (IFR);

(ii) offers infinitely many priority levels (i.e., full separation) if and only if $\frac{1-F(c)}{f(c)}$ is increasing.

Corollary 2.2. *If both $c - \frac{1-F(c)}{f(c)}$ and $\frac{1-F(c)}{f(c)}$ are increasing (e.g., exponential and Weibull distribution), increasing the number of priority levels can increase both the provider's revenue and consumer welfare.*[5] *Thus, full separation is both revenue- and welfare-maximizing.*

Moreover, for the highest type $\bar{c}$, maximizing his utility $U(\bar{c})$ subject to constraints (4)–(8) is equivalent to

$$\max_{s \in \mathrm{MPS}(F)} U(0) + \int_0^{\bar{c}} \frac{s(c) - 1}{f(c)} \,\mathrm{d}F(c). \tag{14}$$

Because $p(0) \geq 0$, (IR) implies $U(0) = p(0) = 0$. By Kleiner, Moldovanu, and Strack (2021, Proposition 2), if f is increasing, $s(c) = 1/2$ maximizes $U(\bar{c})$. In this case, $U(c)$ is linear (because $U'(c) = -1/2$), and thus $U(c) = U(\bar{c}) \cdot c/\bar{c}$ for all $c \in [0, \bar{c}]$. With more priority

[4] This would not be true without the assumption $p(c) \geq 0$. If negative transfers are allowed and subject to budget balance, full separation (assortative matching) will maximize consumer welfare because it is efficient, and transfers can redistribute the efficiency gains among consumers (see, e.g., Gershkov and Schweinzer, 2010). I thank Alex Gershkov for pointing this out.

[5] The Weibull distribution is a distribution on $\mathbb{R}_+$ with CDF $G(x) = 1 - \exp(-x^k)$, where $k \geq 0$ is the shape parameter. When $k = 1$, it is the same as the exponential distribution. When $k < 1$, both $\frac{1-F(x)}{f(x)}$ and $x - \frac{1-F(x)}{f(x)}$ are strictly increasing.

levels, $U(c)$ is convex (while $U(0) = 0$ still holds), so $U(c) \leq U(\bar{c}) \cdot c/\bar{c}$ for all $c \in [0, \bar{c}]$. Therefore, $s(c) = 1/2$ also maximizes $U(c)$ for all $c \in [0, \bar{c}]$.

Proposition 3 (Cf. GW's Proposition 2)**.** *If f is increasing, offering no priority service (total pooling) maximizes every consumer's utility.*

Remark 6. Proposition 3 shows if f is increasing, all consumers are worse off after *any* $k > 1$ levels of PS are offered than if no PS is offered ($k = 1$). GW's Proposition 2 shows that under this assumption, all consumers are worse off after the introduction of one level of PS ($k = 2$) than $k = 1$.

3.3 Regular Service and Priority Service

In GW's Section III, the provider can have at most two priority levels—regular service and priority service. I denote them by $\{p_L, t_L\}$ (regular) and $\{p_H, t_H\}$ (priority), respectively. Under the assumption that the price of the regular service $p_L = 0$, there is no exclusion.

Proposition 4 (Cf. GW's Proposition 1)**.** *The customers' welfare is higher when PS is not offered than when PS is offered if and only if* $\int_0^c \left(\frac{1-F(x)}{f(x)}\right) \mathrm{d}F(x) \geq \mathbf{E}[c]$ *for all* $c \in [0, \bar{c}]$.

Remark 7. GW's Proposition 1 provides a sufficient condition for this result: $\mathbf{E}[c] - \frac{1-F(c)}{f(c)}$ "changes sign at most once, from negative to positive" (i.e., single-crosses zero from below). Because $\int_0^{\bar{c}} \frac{1-F(x)}{f(x)} \mathrm{d}F(c) = \mathbf{E}[c]$, their single-crossing condition (akin to the first-order stochastic dominance) implies the condition in Corollary 4 (akin to the second-order stochastic dominance).

Remark 8. A sufficient condition is the IFR property, according to Proposition 2.

3.4 Discussion on Exclusion

To this point, I have maintained the assumption that the regular service is free so that there is no exclusion. If exclusions are possible, it will affect the feasibility of the waiting time function $t(c)$ by requiring $1 - t$ to be a mean-preserving spread of $1 - F(c)$ (in the quantile space) for included types. For example, if the provider excludes $c \in [0, \hat{c}]$, then $s = 1 - t \in \mathrm{MPS}(F)$ on $[\hat{c}, \bar{c}]$ (in the quantile space).[6] This effect on feasibility is also

[6] In other words, if low types are excluded, the feasibility condition in Theorem 2 becomes: s is *weakly* majorized by F, denoted by $s \in \mathrm{MPS}_w(F)$, that is,

$$\int_x^1 s(c)\,\mathrm{d}F(c) \geq \int_x^1 F(c)\,\mathrm{d}F(c),$$

and the equality need not hold at $x = 0$.

present in the standard monopoly problem (e.g. Mussa and Rosen, 1978), where the revenue-maximizing seller will exclude agents with negative marginal revenue ($J(\theta) < 0$), whereas no agents will be excluded in welfare maximization. However, the participation constraint is trickier here because the consumer's reservation utility, which equals the value of the service to him, is ambiguous.[7] This calls for introducing the value of the service to the model, as in GW's Section IV.

4 Endogenous Pricing of the Regular Service

In GW's Section IV, they allow for endogenous pricing of the regular service (and hence exclusion) and introduce an additional variable—the value of the service to the consumer. They assume the consumer has a unidimensional type $\theta \in [0, \bar{\theta}]$ that determines their unit cost of waiting, $c(\theta)$, and value of the service, $v(\theta)$. The consumer's utility is $v(\theta) - p - tc(\theta)$. Abusing notations, denote the CDF of type θ by $F(\theta)$ and the density by $f(\theta) > 0$ on the support $[0, \bar{\theta}]$. Denote $U(\theta) = v(\theta) - p(\theta) - t(\theta)c(\theta)$.

I maintain their assumptions that $v(\theta) > 0$, $c(\theta) > 0$, $v'(\theta) \geq 0$, $c'(\theta) \geq 0$, and $v'(\theta) > c'(\theta)$ (for low-type exclusion in their Section IV.A).[8]

Lemma 3. *A direct mechanism $\{p(\theta), t(\theta)\}$ is incentive-compatible if and only if*

- *there exists a cutoff type $\hat{\theta} \in [0, \bar{\theta}]$ such that $U(\theta) \geq 0$ if and only if $\theta \geq \hat{\theta}$,*
- $U(\theta) = U(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} (v'(x) - t(x)c'(x))\, \mathrm{d}x$ *for all $\theta \in [\hat{\theta}, \bar{\theta}]$, and*
- $t(\theta)$ *is decreasing.*

Note that (IC) implies $U'(\theta) = v'(\theta) - t(\theta)c'(\theta) > 0$ because $v'(\theta) > c'(\theta)$ and $t(\theta) \in [0, 1]$, so only low types will excluded.

4.1 Revenue maximization

The revenue maximization problem is given by

$$\max_{t(\theta), p(\theta), \hat{\theta}} \int_{\hat{\theta}}^{\bar{\theta}} p(\theta)\, \mathrm{d}F(\theta) \tag{15}$$

[7] In Moldovanu, Sela, and Shi (2007), the reservation utility is zero due to normalization.

[8] In Section IV.B, they assume $v'(\theta) < c'(\theta)$, which leads to high-type exclusion.

subject to the following constraints on $[\hat{\theta}, \bar{\theta}]$:

$$p(\theta) \geq 0 \tag{16}$$

$$U(\theta) \equiv v(\theta) - p(\theta) - t(\theta)c(\theta) \geq 0 \qquad \text{(IR)} \tag{17}$$

$$v(\theta) - p(\theta) - t(\theta)c(\theta) = U(\hat{\theta}) + \int_{\hat{\theta}}^{\theta} (v'(x) - t(x)c'(x))\,\mathrm{d}x \qquad \text{(IC)} \tag{18}$$

$$t(\theta) \text{ is decreasing} \tag{19}$$

$$s \equiv 1 - t \in \mathrm{MPS}(F) \text{ on } [\hat{\theta}, \bar{\theta}] \qquad \text{(MPSw)} \tag{20}$$

Denote $J_c(\theta) = c(\theta) - \frac{1-F(\theta)}{f(\theta)}c'(\theta)$, which equals $J(\theta)$ if $c(\theta) = \theta$ and is increasing if F satisfies the IFR property. Analogous to the previous section, for any given cutoff $\hat{\theta} \in [0, \bar{\theta}]$, the revenue maximization problem is equivalent to

$$\max_{s \in \mathrm{MPS}(F) \text{ on } [\hat{\theta},\bar{\theta}]} \int_{\hat{\theta}}^{\bar{\theta}} J_c(\theta)s(\theta)\,\mathrm{d}F(\theta) + \int_{\hat{\theta}}^{\bar{\theta}} \left[(v(\theta) - c(\theta)) - \frac{1-F(\theta)}{f(\theta)}(v'(\theta) - c'(\theta)) - U(\hat{\theta})\right] \mathrm{d}F(\theta) \tag{21}$$

where $s(\theta) = 1 - t(\theta)$ is increasing. Because $U(\hat{\theta}) \geq 0$ (IR), it is optimal to set $U(\hat{\theta}) = 0$. Because IFR implies that $J_c(\theta)$ is increasing, by Kleiner, Moldovanu, and Strack (2021, Proposition 2), the solution is $t(\theta) = 1 - F(\theta)$ (i.e., $s(\theta) = F(\theta)$). Now we solve for the optimal cutoff $\hat{\theta}$.

$$\max_{\hat{\theta} \in [0,\bar{\theta}]} R(\hat{\theta}) = \max_{\hat{\theta} \in [0,\bar{\theta}]} \int_{\hat{\theta}}^{\bar{\theta}} \left[J_c(\theta)F(\theta) + (v(\theta) - c(\theta)) - \frac{1-F(\theta)}{f(\theta)}(v'(\theta) - c'(\theta))\right] \mathrm{d}F(\theta) \tag{22}$$

Denote $J_{vc}(\theta) = (v(\theta) - c(\theta)) - \frac{1-F(\theta)}{f(\theta)}(v'(\theta) - c'(\theta))$. The FOC is

$$R'(\hat{\theta}) = -(J_c(\hat{\theta})F(\hat{\theta}) + J_{vc}(\hat{\theta}))f(\hat{\theta}) \leq 0, \quad R'(\hat{\theta}) \cdot \hat{\theta} = 0 \tag{23}$$

IFR implies that J_c and J_{vc} are increasing, so R is concave if f is also increasing.

Proposition 5. *Assume F satisfies the IFR property and R is concave. The revenue-maximizing mechanism excludes types below $\hat{\theta}^*$ (given by equation (23)) and offers infinitely many priority levels to types above $\hat{\theta}^*$. In particular, it has no exclusion ($\hat{\theta}^* = 0$) if and only if $v(0) - c(0) - \frac{v'(0)-c'(0)}{f(0)} \geq 0$.*

4.2 Consumer welfare maximization

The consumer welfare maximization problem is given by

$$\max_{t(\theta),p(\theta),\hat{\theta}} \int_{\hat{\theta}}^{\bar{\theta}} U(\theta)\,\mathrm{d}F(\theta) \tag{24}$$

subject to constraints (16)–(20) on $[\hat{\theta}, \bar{\theta}]$.

Analogous to the previous section, for any given cutoff $\hat{\theta} \in [0, \bar{\theta}]$, the welfare maximization problem is equivalent to

$$\max_{s\in\mathrm{MPS}(F)\text{ on }[\hat{\theta},\bar{\theta}]} \int_{\hat{\theta}}^{\bar{\theta}} \left(\frac{1-F(\theta)}{f(\theta)}\right) c'(\theta)s(\theta)\,\mathrm{d}F(\theta) + \int_{\hat{\theta}}^{\bar{\theta}} \left(\frac{1-F(\theta)}{f(\theta)}(v'(\theta)-c'(\theta)) + U(\hat{\theta})\right)\mathrm{d}F(\theta) \tag{25}$$

Assume $\frac{1-F(\theta)}{f(\theta)}c'(\theta)$ is decreasing. Then, the optimal $t(\theta)$ pools all types $\theta \in [\hat{\theta}, \bar{\theta}]$, that is,

$$t(\theta) = \int_{\hat{\theta}}^{\bar{\theta}} 1 - F(\theta)\,\mathrm{d}F(\theta) = 1/2 + F(\hat{\theta})^2/2 - F(\hat{\theta}). \tag{26}$$

Now we solve for the optimal cutoff $\hat{\theta}$.

$$\max_{\hat{\theta}\in[0,\bar{\theta}]} W(\hat{\theta}) = \max_{\hat{\theta}\in[0,\bar{\theta}]} \int_{\hat{\theta}}^{\bar{\theta}} \frac{1-F(\theta)}{f(\theta)}[v'(\theta) - c'(\theta)/2 + c'(\theta)(F(\hat{\theta}) - F(\hat{\theta})^2/2)] + U(\hat{\theta})\,\mathrm{d}F(\theta) \tag{27}$$

Note that $W(\hat{\theta})$ is discontinuous at $\hat{\theta} = 0$. At $\hat{\theta} = 0$, it is optimal to set $p(0) = 0$ and $U(0) = v(0) - t(0)c(0) = v(0) - c(0)/2$, so

$$W(0) = \int_0^{\bar{\theta}} \frac{1-F(\theta)}{f(\theta)}(v'(\theta) - c'(\theta)/2)\,\mathrm{d}F(\theta) + (v(0) - c(0)/2). \tag{28}$$

For $\hat{\theta} > 0$, $U(\hat{\theta}) = 0$, so

$$\begin{aligned} W'(\hat{\theta}) = &-(1-F(\hat{\theta}))(v'(\hat{\theta}) - c'(\hat{\theta})/2) + (1-F(\hat{\theta}))c'(\hat{\theta})(F(\hat{\theta}) - F(\hat{\theta})^2/2) \\ &+ \int_{\hat{\theta}}^{\bar{\theta}} \frac{1-F(\theta)}{f(\theta)} c'(\theta)(1-F(\hat{\theta}))f(\hat{\theta})\,\mathrm{d}F(\theta) \end{aligned} \tag{29}$$

and we need to compare $W(0)$ and $\sup_{\hat{\theta}\in(0,\bar{\theta}]} W(\hat{\theta})$. If $W(0) \geq \sup_{\hat{\theta}\in(0,\bar{\theta}]} W(\hat{\theta})$, then $\hat{\theta}^* = 0$ is optimal. Otherwise, some exclusion $\hat{\theta}^* > 0$ is optimal (assume $W(\theta)$ is concave; then $\hat{\theta}^*$ is the solution to $W'(\hat{\theta}) = 0$ in equation (29)).

Proposition 6. *Assume $\frac{1-F(\theta)}{f(\theta)}c'(\theta)$ is decreasing. The consumer welfare-maximizing mech-*

anism excludes types below $\hat{\theta}^*$ and offers one priority level to types above $\hat{\theta}^*$.

5 Discussions

With a feasibility condition à la Theorem 2, the mechanism design approach can be applied to allocation problems in which allocating the object to one agent has externalities on other agents, such as waiting time (Gershkov and Winter, 2023), status (Moldovanu, Sela, and Shi, 2007), and conspicuous goods (Rayo, 2013).

By applying this approach to GW's model, I study the effect of PS on the monopoly provider's revenue and consumer welfare and provide the necessary and sufficient conditions for some of their propositions while allowing for stochastic priority levels. In particular, I show that the trade-off between the provider's revenue and consumer welfare is less stark. Although an increasing (decreasing) failure rate implies that increasing the number of priority levels increases the revenue but decreases (increases) consumer welfare at the same time, it is not a necessary condition for the former. Thus, for some cost distributions with decreasing failure rates (e.g., exponential and Weibull distributions), the two objectives are aligned—increasing the number of priority levels will increase *both* the revenue and consumer welfare. Moreover, an all-pay auction among customers can implement the mechanism with infinitely many priority levels.

References

Border, Kim C. 1991. "Implementation of Reduced Form Auctions: A Geometric Approach." *Econometrica* 59 (4):1175–1187. [3]

Gershkov, Alex and Paul Schweinzer. 2010. "When Queueing Is Better than Push and Shove." *International Journal of Game Theory* 39 (3):409–430. [6]

Gershkov, Alex and Eyal Winter. 2023. "Gainers and Losers in Priority Services." *Journal of Political Economy* 131 (11):3103–3155. [1, 11]

Kleiner, Andreas, Benny Moldovanu, and Philipp Strack. 2021. "Extreme Points and Majorization: Economic Applications." *Econometrica* 89 (4):1557–1593. [1, 3, 5, 6, 9]

Maskin, Eric and John Riley. 1984. "Optimal Auctions with Risk Averse Buyers." *Econometrica* 52 (6):1473–1518. [3]

Matthews, Steven A. 1984. "On the Implementability of Reduced Form Auctions." *Econometrica* 52 (6):1519–1522. [3]

Moldovanu, Benny, Aner Sela, and Xianwen Shi. 2007. "Contests for Status." *Journal of Political Economy* 115 (2):338–363. [1, 3, 8, 11]

Mussa, Michael and Sherwin Rosen. 1978. "Monopoly and Product Quality." *Journal of Economic Theory* 18 (2):301–317. [8]

Onuchic, Paula and Debraj Ray. 2023. "Conveying Value via Categories." *Theoretical Economics* 18 (4):1407–1439. [3]

Rayo, Luis. 2013. "Monopolistic Signal Provision." *The B.E. Journal of Theoretical Economics* 13 (1):27–58. [3, 11]

Xiao, Peiran. 2024. "Incentivizing Agents through Ratings." Working paper. [3]