

Online Appendix to “Invite Your Friend and You’ll Move up in Line: Optimal Design of Referral Priority Programs”

Luyi Yang

Appendix A: Proofs

Proof of Proposition 1 The first-best problem relaxes IC constraint (5) in Problem 1 and effectively reduces to: $\max_{\alpha, r} \frac{\Lambda\alpha}{1-r\alpha}$, s.t. $V - rk - \frac{c}{\mu(1-r\alpha) - \Lambda\alpha} \geq 0$. Let $\gamma \geq 0$, be the Lagrangian multiplier for the constraint. Thus, the Lagrangian becomes $\mathcal{L} = \frac{\Lambda\alpha}{1-r\alpha} + \gamma \left(V - rk - \frac{c}{\mu(1-r\alpha) - \Lambda\alpha} \right)$. From the KKT conditions $\frac{\partial \mathcal{L}}{\partial r} = \frac{\Lambda\alpha^2}{(1-r\alpha)^2} + \gamma \left(-k - c \frac{\mu\alpha}{[\mu(1-r\alpha) - \Lambda\alpha]^2} \right) = 0$ and $\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{\Lambda}{(1-r\alpha)^2} + \gamma \left(-c \frac{\mu r + \Lambda}{[\mu(1-r\alpha) - \Lambda\alpha]^2} \right) = 0$, we have $\gamma > 0$. By complementary slackness, $V - rk - \frac{c}{\mu(1-r\alpha) - \Lambda\alpha} = 0$. This implies that $r\alpha W_1 + (1-r\alpha)W_0 = \frac{1}{\mu(1-r\alpha) - \Lambda\alpha}$. Hence, the optimal scheduling policy must always be work-conserving in the first-best problem. From $U \equiv V - rk - \frac{c}{\mu(1-r\alpha) - \Lambda\alpha} = 0$, $\alpha(r) \triangleq \frac{\mu(V-rk) - c}{(V-rk)(\Lambda + \mu r)} \leq 1$. Also, $\partial U / \partial \alpha < 0$ and $\partial U / \partial r < 0$. Thus, $d\alpha / dr = -(\partial U / \partial r) / (\partial U / \partial \alpha) < 0$; α is decreasing in r . The objective function can be expressed as a function of r : $\frac{\Lambda\alpha}{1-r\alpha} = \frac{\Lambda \frac{\mu(V-rk) - c}{(V-rk)(\Lambda + \mu r)}}{1 - r \frac{\mu(V-rk) - c}{(V-rk)(\Lambda + \mu r)}} = \Lambda \frac{\mu(V-rk) - c}{\Lambda(V-rk) + cr}$. We shall show this function is decreasing in r by showing its reciprocal is increasing in r . It suffices to show $\frac{\Lambda(V-rk) + cr}{\mu(V-rk) - c} = \frac{\Lambda}{\mu} + \frac{cr + \Lambda c / \mu}{\mu(V-rk) - c}$ is increasing in r , which is obvious. Therefore, the objective function is decreasing in r . Note that $\alpha(0) = \frac{\mu V - c}{V\Lambda}$. Thus, $\alpha(0) \leq 1$ if and only if $\Lambda \geq \mu - c/V$. Thus, when $\Lambda \geq \mu - c/V$, $r = 0$ and $\alpha = \frac{\mu V - c}{V\Lambda}$. When $\Lambda < \mu - c/V$, $\alpha = 1$ and $r \in (0, 1)$ solves $\alpha(r) = 1$, or $\frac{\mu(V-rk) - c}{(V-rk)(\Lambda + \mu r)} = 1$. Collecting terms gives $\mu - \frac{c}{V-rk} - \mu r = \Lambda$. It is easy to see $\partial r / \partial \Lambda < 0$. Solving for r gives $r = \frac{-\sqrt{4k\mu(c + \Lambda V - \mu V) + (-k\Lambda + k\mu + \mu V)^2 - k\Lambda + k\mu + \mu V}}{2k\mu}$. \square

Proof of Proposition 3 Any mechanism that generate referrals must have $k \leq c\alpha(W_0 - W_1)$, or equivalently, $W_0 \geq k/(c\alpha) + W_1$. Moreover, the OA constraint requires $W_1 \geq 1/(\mu(1-r\alpha) - \Lambda r\alpha^2)$ and the IR constraint requires $V - rk - c[r\alpha W_1 + (1-r\alpha)W_0] \geq 0$. Thus, by substitution, any mechanism that generates referrals must have

$$V - \frac{k}{\alpha} - \frac{c}{\mu(1-r\alpha) - \Lambda r\alpha^2} \geq 0. \quad (\text{A.1})$$

Since $\frac{c}{\mu(1-r\alpha) - \Lambda r\alpha^2} > \frac{c}{\mu}$ and $\frac{k}{\alpha} \geq k$, thus to generate referrals, we must have $V - k - \frac{c}{\mu} > 0$. Therefore, if $V - c/\mu \leq k$, no referrals can be generated. \square

Proof of Proposition 4 In the first-best problem, when $\Lambda < \mu - c/V$, $\alpha^{FB} = 1$ and r^{FB} solves

$$V - r^{FB}k - \frac{c}{\mu(1-r^{FB}) - \Lambda} = 0. \quad (\text{A.2})$$

It is achievable by the second best if and only if W_1 and W_0 satisfy $r^{FB}W_1 + (1-r^{FB})W_0 = \frac{1}{\mu(1-r^{FB}) - \Lambda}$, $k = c[W_0 - W_1]$, $W_1 \geq \frac{1}{\mu(1-r^{FB}) - \Lambda r^{FB}}$. The first equation is due to work-conservation of the first-best solution. The second equation is due to $r^{FB} \in (0, 1)$, i.e., customers should be indifferent to referrals. The inequality is from the OA constraint. Solving the first two equations gives $W_1 = \frac{1}{\mu(1-r^{FB}) - \Lambda} - \frac{k(1-r^{FB})}{c}$, $W_0 = \frac{1}{\mu(1-r^{FB}) - \Lambda} + \frac{kr^{FB}}{c}$; Thus, we need to show

$$\frac{1}{\mu(1-r^{FB}) - \Lambda} - \frac{k(1-r^{FB})}{c} \geq \frac{1}{\mu(1-r^{FB}) - \Lambda r^{FB}} \quad (\text{A.3})$$

if and only if $\Lambda \geq \Lambda^*$ (subject to $\Lambda \geq \mu - c/V$), where $\Lambda^* \in (0, \mu - c/V)$ is uniquely determined. Combining (A.2) and (A.3), it follows that (A.3) holds if and only if

$$V - k - \frac{c}{\mu(1 - r^{FB}) - \Lambda r^{FB}} \geq 0. \quad (\text{A.4})$$

From (A.2), $\Lambda = \mu(1 - r^{FB}) - \frac{c}{V - r^{FB}k}$. Note that Λ is decreasing in r^{FB} . Plugging it into (A.4) gives $V - k - \frac{c}{\mu(1 - r^{FB}) - \Lambda r^{FB}} = V - k - \frac{c}{\mu(1 - r^{FB}) - \mu(1 - r^{FB})r^{FB} + r^{FB}\frac{c}{V - r^{FB}k}} = V - k - \frac{c}{\mu(1 - r^{FB})^2 + \frac{r^{FB}c}{V - r^{FB}k}} \triangleq f(r^{FB})$. Thus, proving the proposition is equivalent to showing that when $V - c/\mu > k$, there exists a unique $r^* \in (0, 1)$ such that $f(r^*) = 0$. Moreover, $f(r) > 0$ if $r \in (0, r^*)$ and $f(r) < 0$ if $r \in (r^*, 1)$. By inspection, $f(1) = 0$; $f(0) = V - k - c/\mu > 0$. $f(r) = V - k - \frac{c(V - rk)}{\mu(1 - r)^2(V - rk) + rc}$. Since $\mu(1 - r)^2(V - rk) + rc > 0$ for $r \in (0, 1)$, it suffices to show $g(r) \triangleq f(r)(\mu(1 - r)^2(V - rk) + rc)$ has this property. Plugging $f(r)$ into $g(r)$ gives $g(r) = (\mu(1 - r)^2(V - rk) + rc)(V - k) - c(V - rk)$. Note that $g(r)$ is a third-degree polynomial function, which has at most three roots. We already know one root $r = 1$. The coefficient for r^3 is $-k\mu(V - k) < 0$. Therefore, $g(r) < 0$ for sufficiently large r . The derivative of $g(r)$ evaluated at $r = 1$ is $g'(1) = cV > 0$. This implies there exists $\varepsilon > 0$ such that $g(1 + \varepsilon) > 0$ and $g(1 - \varepsilon) < 0$. Since $g(0) > 0$ and $g(1 - \varepsilon) < 0$, there exists $r^* \in (0, 1)$ such that $g(r^*) = 0$ by the intermediate value theorem. Likewise, since $g(1 + \varepsilon) > 0$ and $g(r) < 0$ for sufficiently large r , there also exists an $r^{*'} > 1$ such that $g(r^{*'}) = 0$. Since $g(r)$ has at most three roots and one root is 1, we conclude that r^* and $r^{*'}$ must be unique. Moreover, $g(r) > 0$ if $r \in (0, r^*)$ and $g(r) < 0$ if $r \in (r^*, 1)$. Therefore, f also has this property. We also need to show $\Lambda^* = \mu(1 - r^*) - \frac{c}{V - r^*k} > 0$. Since Λ is decreasing in r , It suffices to show $f(r_0) < 0$, where r_0 solves $\mu(1 - r_0) - \frac{c}{V - r_0k} = 0$. By definition, $f(r_0) = V - k - \frac{c}{\mu(1 - r_0)} = V - k - \frac{c}{V - r_0k} = -(1 - r_0)k < 0$. Therefore, $\Lambda^* > 0$. Finally, solving $g(r^*) = 0$ for r^* gives $r^* = \frac{\sqrt{(\mu V^2 - k^2 \mu)^2 - 4(k^2 \mu - k \mu V)(cV + k \mu V - \mu V^2)} + k^2 \mu - \mu V^2}{2(k^2 \mu - k \mu V)}$ and $\Lambda^* = \left[\mu(1 - r^*) - \frac{c}{V - r^*k} \right]$.

Next, we identify the optimal second-best mechanism for $\Lambda < \Lambda^*$ and $k < V - c/\mu$. Either $r = 0$ (FIFO) or $r > 0$ (referrals generated). Now, we compare these two candidate mechanisms.

Case 1: $r = 0$. Since $\Lambda < \Lambda^* < \mu - c/V$, $V - c/(\mu - \Lambda) > 0$; thus, $\alpha = 1$, and the system throughput is Λ .

Case 2: $r > 0$. From the proof of Proposition 3, any mechanism that generates referrals must satisfy (A.1). Thus, the optimal value to the following relaxation problem is an upper bound of the system throughput with $r > 0$: $\max_{\alpha, r} \frac{\Lambda \alpha}{1 - r \alpha}$ s.t. $V - \frac{k}{\alpha} - \frac{c}{\mu(1 - r \alpha) - \Lambda r \alpha^2} \geq 0$. Note that at this point, we have not shown that the solution of this problem is implementable. We will show this later. In this relaxation problem, the constraint must be binding because otherwise one can always increase r to increase the objective function while still satisfying the constraint. Therefore, $V - k/\alpha - \frac{c}{\mu(1 - r \alpha) - \Lambda r \alpha^2} = 0$, which gives $r = \frac{-\alpha c + \alpha \mu V - k \mu}{\alpha(\alpha V - k)(\alpha \Lambda + \mu)}$. Plugging it into the objective function gives $\Lambda \frac{\alpha}{1 - r \alpha} = \Lambda \frac{(\alpha V - k)(\alpha \Lambda + \mu)}{\alpha \Lambda V + c - k \Lambda} = \Lambda \frac{\alpha \Lambda + \mu}{\Lambda + \frac{c}{\alpha V - k}}$. From $r = \frac{-\alpha c + \alpha \mu V - k \mu}{\alpha(\alpha V - k)(\alpha \Lambda + \mu)}$, $\alpha V - k = \frac{c \alpha}{\mu(1 - r \alpha) - \Lambda r \alpha^2} > 0$. Therefore, $\Lambda \alpha / (1 - r \alpha) = \Lambda \frac{\alpha \Lambda + \mu}{\Lambda + \frac{c}{\alpha V - k}}$ is increasing in α . Hence, the optimal $\alpha^{\text{SB}} = 1$. Plugging $\alpha = 1$ into $r = \frac{-\alpha c + \alpha \mu V - k \mu}{\alpha(\alpha V - k)(\alpha \Lambda + \mu)}$ gives $r^{\text{SB}} = \frac{\mu(V - k) - c}{(V - k)(\mu + \Lambda)}$. Note that the throughput in this relaxation problem is obviously higher than the FIFO throughput Λ in Case 1. Since $V - k/\alpha - \frac{c}{\mu(1 - r \alpha) - \Lambda r \alpha^2} = 0$, it follows from the logic of the proof of Proposition 3 that $W_1 = \frac{1}{\mu(1 - r) - \Lambda r}$, $W_0 = W_1 + k/c$. It remains to be shown that the solution to this relaxation problem is implementable. We check if all the constraints in the mechanism design problem are satisfied. The IC constraint is satisfied by construction since $c\alpha(W_0 - W_1) = k$. The IR constraint is satisfied because

$$V - rk - c[r\alpha W_1 - (1 - r\alpha)W_0] = V - k - \frac{c}{\mu(1 - r) - \Lambda r} = 0. \quad (\text{A.5})$$

The OA constraint $W_1 \geq 1/(\mu(1-r\alpha) - \Lambda r\alpha^2)$ is obviously satisfied. We check the other OA constraint: $r\alpha W_1 + (1-r\alpha)W_0 \geq 1/(\mu(1-r\alpha) - \Lambda\alpha)$. From (A.5), it is equivalent to checking

$$V - rk - \frac{c}{\mu(1-r) - \Lambda} \geq 0. \quad (\text{A.6})$$

We shall show that this is true and the inequality is strict if $\Lambda < \Lambda^*$, which would prove the optimality of strategic delay. First note that by definition of Λ^* , at Λ^* , $V - r^*k - \frac{c}{\mu(1-r^*\alpha) - \Lambda^*\alpha} = 0$, where r^* solves $V - k - \frac{c}{\mu(1-r^*) - \Lambda^*r^*} = 0$. From (A.5), $\Lambda = \mu(1-r)/r - \frac{c}{(V-k)r}$. Moreover, decreasing Λ increases r , which implies that for $\Lambda < \Lambda^*$, $r > r^*$. Plugging this into the LHS of (A.6) gives $V - rk - \frac{c}{\mu(1-r) - \Lambda} = V - rk - \frac{c}{\mu(1-r) - \mu(1-r^*)/r + \frac{c}{(V-k)r}} = V - rk - \frac{c(V-k)r}{c - (V-k)\mu(1-r)^2} \triangleq \phi(r)$. To prove (A.6), it suffices to show $\phi(r) > 0$ for $r \in (r^*, 1)$. Note that $c - (V-k)\mu(1-r)^2 > 0$ since $\frac{c(V-k)r}{c - (V-k)\mu(1-r)^2} = \frac{c}{\mu(1-r) - \Lambda} > 0$ and $c(V-k)r > 0$. Therefore, it is equivalent to showing $(V - rk)(c - (V-k)\mu(1-r)^2) - c(V-k)r > 0$ for $r \in (r^*, 1)$. Recognize that the LHS is equal to $-g(r)$, where $g(r)$ is defined in the proof of Proposition 4, in which we show $g(r) < 0$ for $r > r \in (r^*, 1)$. Therefore, $-g(r) > 0$ for $r > r \in (r^*, 1)$. This shows that the solution to the relaxation problem is implementable, and since it achieves a higher throughput than FIFO in Case 1. This (strategic-delay) solution is optimal. \square

Proof of Proposition 5 The pair (Λ^*, r^*) solves the following joint equations:

$$V - r^*k - \frac{c}{\mu(1-r^*) - \Lambda^*} = 0, \quad V - k - \frac{c}{\mu(1-r^*) - \Lambda^*r^*} = 0. \quad (\text{A.7})$$

Expressing Λ^* as a function of Λ from the first equation of (A.7) and plugging it into the second equation of (A.7) gives $g(r^*) = [\mu(1-r^*)^2(V - r^*k) + r^*c](V - k) - cV + cr^*k = 0$. Note that $g(\cdot)$ is the same g function defined in the proof of Proposition 4. Simplifying $g(\cdot)$ yields $g(r) = (1-r)[\mu(1-r)(V - rk)(V - k) - cV] = 0$. Since $r^* \in (0, 1)$, $h(r^*) = \mu(1-r^*)(V - r^*k)(V - k) - cV = 0$. By inspection, $\frac{\partial h(r)}{\partial r} < 0$, $\frac{\partial h}{\partial k} < 0$. Therefore, $\frac{\partial r^*}{\partial k} = -\frac{\partial h/\partial k}{\partial h/\partial r} < 0$. Next, we shall show $\partial\Lambda^*/\partial r^* < 0$, which would prove $\partial\Lambda^*/\partial k > 0$. From the second equation of (A.7), $k = V - c/(\mu(1-r^*) - \Lambda^*r^*)$. Plugging it into the first equation of (A.7) gives $V - r^* \left(V - \frac{c}{\mu(1-r^*) - \Lambda^*r^*} \right) - \frac{c}{\mu(1-r^*) - \Lambda^*} = 0$. Collecting terms gives $V(1-r^*) - \frac{c}{\mu(1-r^*) - \Lambda^*} + \frac{cr^*}{\mu(1-r^*) - \Lambda^*r^*} = 0$. $V(1-r^*) - \frac{c\mu(1-r^*)^2}{(\mu(1-r^*) - \Lambda^*)(\mu(1-r^*) - \Lambda^*r^*)} = 0$. $\tau(\Lambda^*, r^*) = V(1-r^*) - \frac{c\mu}{(\mu - \Lambda^*/(1-r^*))(\mu - \Lambda^*r^*/(1-r^*))} = 0$. By inspection, $\frac{\partial \tau}{\partial r^*} < 0$, $\frac{\partial \tau}{\partial \Lambda^*} < 0$. Therefore, $\frac{\partial \Lambda^*}{\partial r^*} = -\frac{\partial \tau/\partial r^*}{\partial \tau/\partial \Lambda^*} < 0$. Since $\frac{\partial \Lambda^*}{\partial r^*} < 0$ and $\frac{\partial r^*}{\partial k} < 0$, by the chain rule, $\frac{\partial \Lambda^*}{\partial k} = \frac{\partial \Lambda^*}{\partial r^*} \frac{\partial r^*}{\partial k} > 0$. \square

Proof of Proposition 6 The first-best problem can simplify to $\max_{\mathbf{r} \in [0,1]^D, \alpha \in [0,1]} \Pi(\mathbf{r}, \alpha) = \frac{\Lambda\alpha}{1-m(\mathbf{r}, \alpha)}$, s.t. $U(\mathbf{r}, \alpha) = V - k \sum_{i=1}^D r_i - c \frac{\sigma^2(\mathbf{r}, \alpha) + (2-m(\mathbf{r}, \alpha))(1-m(\mathbf{r}, \alpha))}{2(1-m(\mathbf{r}, \alpha))[\mu(1-m(\mathbf{r}, \alpha)) - \Lambda\alpha]} \geq 0$, where $m(\mathbf{r}, \alpha) = \sum_{i=1}^D r_i\alpha$, and $\sigma^2(\mathbf{r}, \alpha) = \sum_{i=1}^D r_i\alpha(1-r_i\alpha)$. To prove Proposition 6, we need to prove that any feasible solution to the first-best problem (\mathbf{r}^*, α^*) with two or more positive r_i^* 's can be strictly dominated by a feasible solution with at most one positive $r_i > 0$. Let $m^* = \sum_{i=1}^D r_i^*\alpha^*$. Then, $U(\mathbf{r}^*, \alpha^*) = V - k \sum_{i=1}^D r_i^* - c \frac{\sum_{i=1}^D r_i^*\alpha^*(1-r_i^*\alpha^*) + (2-m^*)(1-m^*)}{2(1-m^*)[\mu(1-m^*) - \Lambda\alpha^*]} \geq 0$. We first prove the following lemma:

LEMMA A.1. *For any $K \in (0, 1)$, the minimization problem $\min_{x_1, \dots, x_D \in [0, 1], x_1 \geq x_2 \geq \dots \geq x_D} \sum_{i=1}^D x_i(1-x_i)$ s.t. $\sum_{i=1}^D x_i = m$ has the optimal value $m(1-m)$ with the unique optimal solution $x_1 = m$ and $x_2 = x_3 = \dots = x_D = 0$.*

Proof of Lemma A.1 To see this, note $\sum_{i=1}^D x_i(1-x_i) = \sum_{i=1}^D x_i - \sum_{i=1}^D x_i^2 = m - (\sum_{i=1}^D x_i)^2 + 2\sum_{i,j:i \neq j} x_i x_j = m - m^2 + 2\sum_{i,j:i \neq j} x_i x_j \geq m(1-m)$. On the other hand, $x_1 = m$ and $x_2 = x_3 = \dots = x_D = 0$ gives $\sum_{i=1}^D x_i(1-x_i) = m(1-m)$. Moreover, $2\sum_{i,j:i \neq j} x_i x_j > 0$ if there are two positive x_i 's, which implies any solution with two positive x_i 's cannot achieve the objective value $m(1-m)$. Hence, $x_1 = m$ and $x_2 = x_3 = \dots = x_D = 0$ is the unique optimal solution and $m(1-m)$, the optimal value. \square

Lemma A.1 implies that for any (\mathbf{r}^*, α^*) with two or more positive r_i^* 's, $\frac{\sum_{i=1}^D r_i^* \alpha^* (1-r_i^* \alpha^*) + (2-m^*)(1-m^*)}{2(1-m^*)[\mu(1-m^*) - \Lambda \alpha^*]} > \frac{m^*(1-m^*) + (2-m^*)(1-m^*)}{2(1-m^*)[\mu(1-m^*) - \Lambda \alpha^*]} = \frac{1}{\mu(1-m^*) - \Lambda \alpha^*}$. Moreover, since $U(\mathbf{r}^*, \alpha^*) \geq 0$, we have

$$V - k \frac{m^*}{\alpha^*} - \frac{c}{\mu(1-m^*) - \Lambda \alpha^*} > 0. \quad (\text{A.8})$$

Now, we discuss two cases.

Case 1: $\alpha^*/(1-m^*) < 1$. Let $\alpha' = \alpha^*/(1-m^*)$. Thus, $U_1(\alpha') \equiv V - \frac{c}{\mu - \Lambda \alpha'} = V - \frac{c(1-m^*)}{\mu(1-m^*) - \Lambda \alpha^*} > V - \frac{c}{\mu(1-m^*) - \Lambda \alpha^*} > 0$, where the last inequality follows from (A.8). Since $U_1(\alpha')$ is decreasing in α' , there must exist a unique solution $\alpha'' > \alpha'$ such that $U_1(\alpha'') = 0$. Let $\alpha''' = \min\{\alpha'', 1\}$. Thus, $\Lambda \alpha''' > \Lambda \alpha^*/(1-m^*)$. This implies that a non-referral solution $(\mathbf{r}, \alpha) = (\mathbf{0}, \alpha''')$ can strictly dominate (\mathbf{r}^*, α^*) with two or more positive r_i^* 's, and therefore, the latter is not an optimal solution to the first-best problem.

Case 2: $\alpha^*/(1-m^*) \geq 1$. Let $r'_1 = 1 - (1-m^*)/\alpha^* \in [0, 1)$. $U_2(r'_1) = V - k r'_1 - \frac{c}{\mu(1-r'_1) - \Lambda} = V - k(1 - \frac{1}{\alpha^*} + \frac{m^*}{\alpha^*}) - \frac{c \alpha^*}{\mu(1-m^*) - \Lambda \alpha^*} > V - k \frac{m^*}{\alpha^*} - \frac{c}{\mu(1-m^*) - \Lambda \alpha^*} > 0$, where the last inequality follows from (A.8). Since $U_2(r'_1)$ is decreasing in r'_1 , there must exist a unique solution $r''_1 \in (r'_1, 1)$ such that $U_2(r''_1) = 0$. Note that $\Lambda/(1-r''_1) > \Lambda/(1-r'_1) = \Lambda \alpha^*/(1-m^*)$. This implies that a new solution (\mathbf{r}, α) with $r_1 = r''_1$, $r_2 = r_3 = \dots = r_D = 0$ and $\alpha = 1$ can strictly dominate (\mathbf{r}^*, α^*) with two or more positive r_i^* 's, and therefore, the latter is not an optimal solution to the first-best problem. Combining the two cases shows that the first-best problem can have at most one positive r_i in the optimal solution. \square

Proof of Proposition 7 Under the transfer mechanism proposed, we show that (α^{FB}, r^{FB}) satisfies both IR and IC and thus is an equilibrium outcome. If customers adopt strategy (α, r) , then each customer's expected utility is $U(r, \alpha) = \alpha\{V - rk + r\alpha P - c[r\alpha W_1(r, \alpha) + (1-r\alpha)W_0(r, \alpha)] - r^{FB}P\}$, where $W_1(r, \alpha) = 1/[\mu(1-r\alpha) - \Lambda r\alpha^2]$ and $r\alpha W_1(r, \alpha) + (1-r\alpha)W_0(r, \alpha) = 1/(\mu(1-r\alpha) - \Lambda \alpha)$. Thus, $U(r^{FB}, \alpha^{FB} = 1) = V - r^{FB}k - \frac{c}{\mu(1-r^{FB}) - \Lambda} = 0$. Hence, the IR constraint is satisfied by (α^{FB}, r^{FB}) . To satisfy the IC constraint, we must have $k = c\alpha[W_0(r^{FB}, \alpha^{FB}) - W_1(r^{FB}, \alpha^{FB})] + P$, which is satisfied by construction. Thus, the proposed transfer mechanism can induce the first-best customer outcome. Also, note that it is budget-balanced and therefore, it achieves the first-best system throughput without any (long-run) financial cost/gain for the firm. Note that when the first-best is not achieved in the base model, we must have $V - k - c/[\mu(1-r^{FB}) - \Lambda r^{FB}] < 0$; since $V - r^{FB}k - \frac{c}{\mu(1-r^{FB}) - \Lambda} = 0$, we must have $P = k - c \frac{\Lambda}{[\mu(1-r^{FB}) - \Lambda][\mu(1-r^{FB}) - \Lambda r^{FB}]} > 0$. \square

Proof of Proposition 8 We first solve for the optimal FIFO capacity. Given capacity μ , the expected utility from joining a FIFO queue is $V - c/(\mu - \Lambda \alpha)$. Thus, the system throughput $\lambda^{FIFO}(\mu)$ is: $\lambda^{FIFO}(\mu) = 0$, if $\mu < c/V$; $\lambda^{FIFO}(\mu) = \mu - c/V$, if $c/V \leq \mu \leq \Lambda + c/V$; $\lambda^{FIFO}(\mu) = \Lambda$, if $\mu > \Lambda + c/V$. The optimal objective value is $\Pi^{FIFO} = [\max_{\mu} \lambda^{FIFO}(\mu) - \omega \mu^2 / 2]^+$. Since $V > c/\Lambda$, we have $\Pi^{FIFO} = \Lambda - \frac{1}{2}\omega(\Lambda + c/V)^2$, if $\omega < V/(\Lambda V + c)$; $\Pi^{FIFO} = -c/V + \frac{1}{2\omega}$, if $\omega \in [V/(\Lambda V + c), V/(2c)]$; $\Pi^{FIFO} = 0$, if $\omega \geq V/(2c)$. The corresponding optimal

FIFO capacity μ^{FIFO} is $\mu^{FIFO} = \Lambda + c/V$, if $w < \frac{V}{\Lambda V + c}$; $\mu^{FIFO} = 1/\omega$, if $w \in [\frac{V}{\Lambda V + c}, \frac{V}{2c}]$; $\mu^{FIFO} = 0$, if $w \geq \frac{V}{2c}$. Next, we characterize the first best of the referral program. By Proposition 1, in the first best, the system throughput as a function of capacity μ , $\lambda^{FB}(\mu)$, is $\lambda^{FB}(\mu) = 0$, if $\mu < c/V$; $\lambda^{FB}(\mu) = \mu - c/V$, if $c/V \leq \mu \leq \Lambda + c/V$; $\lambda^{FB}(\mu) = \Lambda/(1 - r^{FB}(\mu))$, if $\mu > \Lambda + c/V$; where $r^{FB}(\mu)$ solves $V - kr^{FB}(\mu) - \frac{c}{\mu(1-r^{FB}(\mu))-\Lambda} = 0$. Note that $r^{FB}(\mu)$ is increasing in μ . Thus, $\lambda^{FB}(\mu) = \Lambda/[1 - r^{FB}(\mu)]$ is increasing in μ . Next, we show that $\lambda^{FB}(\mu)$ is concave in μ for $\mu > \Lambda + c/V$. Plugging $r^{FB}(\mu) = 1 - \Lambda/\lambda^{FB}$ into $V - kr^{FB}(\mu) - \frac{c}{\mu(1-r^{FB}(\mu))-\Lambda} = 0$ gives $V - k + k\frac{\Lambda}{\lambda^{FB}} - \frac{c\lambda^{FB}}{\Lambda(\mu - \lambda^{FB})} = 0$. Writing μ as a function of λ^{FB} gives $\mu = \frac{c\lambda^{FB}}{\Lambda[V - k + k\frac{\Lambda}{\lambda^{FB}}]} + \lambda^{FB}$. Note that $V - k + k\frac{\Lambda}{\lambda^{FB}} > 0$, which is equivalent to $\lambda^{FB}V - \lambda^{FB}k + k\Lambda > 0$.

$$\frac{d\mu}{d\lambda^{FB}} = 1 + \frac{c\lambda^{FB}(\lambda^{FB}V - k\lambda^{FB} + 2k\Lambda)}{\Lambda(k(-\lambda^{FB} + \Lambda) + \lambda^{FB}V)^2} > 1; \quad \frac{d^2\mu}{d(\lambda^{FB})^2} = \frac{2ck^2\Lambda}{(\lambda^{FB}V - k\lambda^{FB} + k\Lambda)^3} > 0. \quad (\text{A.9})$$

Therefore, μ is convex increasing in λ^{FB} . From Mršević (2008), the inverse of a convex increasing function is a concave increasing function, i.e., λ^{FB} is concave increasing in μ . Since the capacity cost is $\omega\mu^2/2$ per unit time, the first order condition of the objective function is $\Pi'(\mu) = d\lambda^{FB}(\mu)/d\mu - \omega\mu$. Since λ^{FB} is concave in μ for $\mu > \Lambda + c/V$, $\Pi'(\mu)$ is decreasing in μ for $\mu > \Lambda + c/V$. It follows from the equation $\mu = \frac{c\lambda^{FB}}{\Lambda[V - k + k\frac{\Lambda}{\lambda^{FB}}]} + \lambda^{FB}$ that when $\mu = \Lambda + c/V$, $\lambda^{FB} = \Lambda$. From (A.9), $\frac{d\mu}{d\lambda^{FB}}|_{\lambda^{FB}=\Lambda} = 1 + \frac{c(V+k)}{\Lambda V^2}$. Hence, $\frac{d\lambda^{FB}}{d\mu}|_{\mu=\Lambda+c/V} = \frac{1}{1 + \frac{c(V+k)}{\Lambda V^2}} = \frac{\Lambda V^2}{\Lambda V^2 + c(V+k)}$. Therefore, $\Pi'(\Lambda + c/V) = \frac{\Lambda V^2}{\Lambda V^2 + c(V+k)} - \omega(\Lambda + c/V)$. If $\frac{\Lambda V^2}{\Lambda V^2 + c(V+k)} - \omega(\Lambda + c/V) > 0$, i.e., $\omega < \frac{\Lambda V^2}{[\Lambda V^2 + c(V+k)](\Lambda + c/V)} \triangleq \bar{\omega}$, then there exists a unique $\mu_0 > \Lambda + c/V$ such that $\Pi'(\mu_0) = 0$. Also, From (A.9), $d\mu/d\lambda^{FB} > 1$ implies $d\lambda^{FB}/d\mu < 1$. It follows from $d\lambda^{FB}/d\mu|_{\mu=\mu_0} - \omega\mu_0 = 0$ that $\mu_0 < 1/\omega$. If $\omega \geq \bar{\omega}$, then $\mu \leq \Lambda + c/V$ and the optimal capacity follows from the FIFO case. Thus, $\mu^{FB} = \mu_0 > \Lambda + c/V$ (with $\mu^{FIFO} = \Lambda + c/V$) if $\omega < \bar{\omega}$ and $\mu^{FB} = \mu^{FIFO}$ otherwise. This completes the proof. \square

Proof of Proposition 9 If $\omega \in [\bar{\omega}, V/(2c)]$, the first best is trivially achieved by non-referral FIFO. Next, consider $\omega < \bar{\omega}$. If $k \geq V - c/\mu_0$, then under μ_0 , referrals cannot be generated (by Proposition 3). Thus, the firm runs a FIFO queue with $\mu^{SB} = \Lambda + c/V < \mu_0$. If $k \geq V$, then obviously $k \geq V - c/\mu_0$ and then non-referral FIFO is optimal. This proves Case (iv) of the Proposition.

Next, we consider $k < V$. By Proposition 4, a referral mechanism must satisfy: $\max_{\mu, r} \frac{\Lambda}{1-r} - \frac{1}{2}\omega\mu^2$,

$$\text{s.t. } V - rk - \frac{c}{\mu(1-r) - \Lambda} \geq 0, \quad (\text{A.10})$$

$$V - k - \frac{c}{\mu(1-r) - \Lambda r} \geq 0, \quad (\text{A.11})$$

whereas a FIFO mechanism (with all customers joining) has the optimal objective function value $\Pi^{FIFO} = \Lambda - \frac{1}{2}\omega(\Lambda + c/V)^2$. Note that since $\omega < \bar{\omega}$, the optimal FIFO mechanism must have all customers join the queue. Also note that in the referral mechanism formulation, both Constraints (A.10) and (A.11) being non-binding is not optimal because one can always increase r (without changing μ) to satisfy both constraints to strictly increase the objective function value. Thus, the referral mechanism can have three possibilities at optimality. (1) Both Constraints (A.10) and (A.11) are binding (which corresponds to full priority and it achieves the first best, which dominates the FIFO mechanism); (2) Constraint (A.10) is binding, but not Constraint (A.11) (which corresponds to partial priority and it achieves the first best, which dominates the FIFO mechanism); (3) Constraint (A.11) is binding, but not Constraint (A.10) (which corresponds to

strategic delay; it does not achieve the first best and we must also compare the optimal objective value with the FIFO mechanism to decide which one is better). In the first best (μ^{FB}, r^{FB}) , since $w < \bar{\omega}$, $r^{FB} > 0$ and $V - r^{FB}k - \frac{c}{\mu^{FB}(1-r^{FB})-\Lambda} = 0$. We examine conditions under which (μ^{FB}, r^{FB}) satisfies (A.11) with $\mu = \mu^{FB}$ and $r = r^{FB}$. Note that as ω decreases, μ^{FB} increases and so does r^{FB} . From $V - r^{FB}k - \frac{c}{\mu^{FB}(1-r^{FB})-\Lambda} = 0$ we have $\mu^{FB}(1-r^{FB}) = \Lambda + \frac{c}{V - kr^{FB}}$. Plugging it into the following gives: $V - k - \frac{c}{\mu^{FB}(1-r^{FB})-\Lambda r^{FB}} = V - k - \frac{c}{\Lambda + c/(V - kr^{FB}) - \Lambda r^{FB}} \triangleq \bar{h}(r^{FB})$. Recognize that $\bar{h}(1) = 0$. Also, let $h(r) \triangleq c/(V - kr) - \Lambda r$. Thus, $\bar{h}(r) = V - k - \frac{c}{\Lambda + h(r)}$, $h'(r) = \frac{ck}{(V - kr)^2} - \Lambda$. By inspection, $h'(r)$ is increasing in r . Therefore, $h(r)$ is convex in $r \in (0, 1)$, and since $\bar{h}(r)$ is increasing in $h(r)$, $\bar{h}(r)$ is quasi-convex in $r \in (0, 1)$. Since $\bar{h}(r)$ is quasi-convex and $\bar{h}(1) = 0$, we have the following three possibilities regarding the signs of $\bar{h}(0)$ and $\bar{h}'(1)$: Case 1: $\bar{h}(0) \geq 0$ and $\bar{h}'(1) \leq 0$. Case 2: $\bar{h}(0) \leq 0$ and $\bar{h}'(1) \geq 0$. Case 3: $\bar{h}(0) > 0$ and $\bar{h}'(1) > 0$.

Case 1: $\bar{h}(0) \geq 0$ and $\bar{h}'(1) \leq 0$. In this case, $\bar{h}(r) \geq 0, \forall r \in (0, 1)$. Further, $\bar{h}'(1) \leq 0$ is equivalent to $h'(1) \leq 0$. Thus, $\bar{h}(0) = V - k - \frac{c}{\Lambda + c/V} \geq 0$, and $\bar{h}'(1) = \frac{ck}{(V-k)^2} - \Lambda \leq 0$. These conditions are equivalent to $\Lambda \geq \frac{ck}{(V-k)^2} \triangleq \bar{\Lambda}$. Hence, if $\Lambda \geq \bar{\Lambda}$, $\bar{h}(r) > 0$ for all any $r \in (0, 1)$, which implies the first best can be achieved by partial priority for all $\omega < \bar{\omega}$. This proves Case (i) of the Proposition.

Case 2: $\bar{h}(0) \leq 0$ and $\bar{h}'(1) \geq 0$. In this case, $\bar{h}(r) < 0, \forall r \in (0, 1)$. Conditions $\bar{h}(0) \leq 0$ and $\bar{h}'(1) \geq 0$ is equivalent to

$$V - k - \frac{c}{\Lambda + c/V} \leq 0. \quad (\text{A.12})$$

That is, $\Lambda \leq \frac{ck}{(V-k)V} \triangleq \underline{\Lambda}$. In this case, the first best cannot be achieved for any $w < \bar{\omega}$. Thus, the second best (μ^{SB}, r^{SB}) is either FIFO, or a referral mechanism with strategic delay that satisfies $\chi(\mu^{SB}, r^{SB}) \triangleq V - k - \frac{c}{\mu(1-r^{SB})-\Lambda r^{SB}} = 0$, $V - r^{SB}k - \frac{c}{\mu(1-r^{SB})-\Lambda} > 0$. Note that when $\mu = \Lambda + c/V$ and $r = 0$, $\chi(\mu, r) = V - k - \frac{c}{\Lambda + c/V} < 0$. Also, $\chi(\mu, r)$ is increasing in μ and decreasing in r . Hence, any (μ, r) that satisfies $\chi(\mu, r) = 0$ must have $\mu > \Lambda + c/V = \mu^{FIFO}$. This implies that in this case, $\mu^{SB} > \mu^{FIFO}$. From $\chi(\mu, r) = 0$, we have

$$\mu = \frac{\Lambda r + c/(V-k)}{1-r} = \frac{\Lambda + \Lambda(r-1) + c/(V-k)}{1-r} = -\Lambda + \frac{\Lambda + c/(V-k)}{1-r}. \quad (\text{A.13})$$

Let $x = \frac{\Lambda}{1-r}$. $\frac{\Lambda}{1-r} - \frac{1}{2}\omega\mu^2 = x - \frac{\omega}{2} \left[-\Lambda + x \left(1 + \frac{c}{\Lambda(V-k)} \right) \right]^2 \triangleq \Pi(x)$. $\Pi(x)$ is a concave quadratic function. The first-order condition is $\Pi'(x) = 1 + \Lambda\omega \left(1 + \frac{c}{\Lambda(V-k)} \right) - \omega x \left(1 + \frac{c}{\Lambda(V-k)} \right)^2$, $x \geq \Lambda$. $\Pi'(x)$ is decreasing in x and $\Pi'(\infty) = -\infty$. Thus, if $\Pi'(\Lambda) > 0$ then Π is maximized at x^* , where x^* is the unique solution to $\Pi'(x^*) = 0$; Otherwise, $\Pi(x)$ is decreasing in x , and the maximizer $x^* = \Lambda$ (i.e., no referrals). Also, $\Pi(\Lambda) = \Lambda - \frac{\omega}{2} \left(\frac{c}{V-k} \right)^2 < \Lambda - \frac{\omega}{2} \left(\Lambda + \frac{c}{V} \right)^2 = \Pi^{FIFO}$, where the inequality follows from (A.12). This implies that when $x^* = \Lambda$, the optimal second-best mechanism is non-referral FIFO. $\Pi'(\Lambda) = 1 - \omega \left(1 + \frac{c}{\Lambda(V-k)} \right) \frac{c}{(V-k)}$. $\Pi'(\Lambda)$ is decreasing in ω ; thus, when ω is small enough, $\Pi'(\Lambda) > 0$ and there exists a unique $x^* > 0$ such that $\Pi'(x^*) = 0$. We next show that when $\omega = \bar{\omega}$, $\Pi'(\Lambda) < 0$. $\Pi'(\Lambda)|_{\omega=\bar{\omega}} = 1 - \bar{\omega} \left(1 + \frac{c}{\Lambda(V-k)} \right) \frac{c}{(V-k)} = 1 - \frac{\Lambda V^2}{[\Lambda V^2 + c(V+k)](\Lambda + c/V)} \left(1 + \frac{c}{\Lambda(V-k)} \right) \frac{c}{(V-k)}$. From (A.12), $c/(V-k) > \Lambda + c/V$. Hence, $1 - \frac{\Lambda V^2}{[\Lambda V^2 + c(V+k)](\Lambda + c/V)} \left(1 + \frac{c}{\Lambda(V-k)} \right) \frac{c}{(V-k)} < 1 - \frac{\Lambda V^2}{[\Lambda V^2 + c(V+k)](\Lambda + c/V)} \left(1 + \frac{c}{\Lambda(V-k)} \right) (\Lambda + c/V) = 1 - \frac{V^2[\Lambda(V-k)+c]}{[\Lambda V^2 + c(V+k)](V-k)} = 1 - \frac{\Lambda V^2(V-k)+cV^2}{\Lambda V^2(V-k)+c(V^2-k^2)} < 0$. Letting $\Pi'(\Lambda) = 0$ gives $w = \frac{\Lambda(V-k)^2}{c[\Lambda(V-k)+c]} \triangleq \tilde{\omega}$. Therefore, for $\omega \in [\tilde{\omega}, \bar{\omega})$, the optimal second-best mechanism is non-referral FIFO. When $\omega < \tilde{\omega}$, the referral mechanism

with strategic delay (SD) is feasible. Its system throughput is $\lambda^{SD} = \frac{1+\Lambda\omega(1+\frac{c}{\Lambda(V-k)})}{\omega(1+\frac{c}{\Lambda(V-k)})^2}$. The optimal capacity of the referral mechanism is $\mu^{SD} \triangleq \frac{\Lambda(V-k)}{\omega[\Lambda(V-k)+c]}$. The optimal profit under that mechanism is $\Pi^{SD} = \frac{\Lambda^2(V-k)(V-k+2(c+\Lambda(V-k))\omega)}{2(c+\Lambda(V-k))^2\omega} = \frac{\Lambda^2(V-k)^2}{2(c+\Lambda(V-k))^2\omega} + \frac{\Lambda^2(V-k)}{(c+\Lambda(V-k))}$. The difference between the referral profit and the FIFO profit $\Pi^{FIFO} = \Lambda - \frac{1}{2}\omega(\Lambda + c/V)^2$ is

$$\Pi^{SD} - \Pi^{FIFO} = \frac{\Lambda^2(V-k)^2}{2(c+\Lambda(V-k))^2\omega} + \frac{(\Lambda + c/V)^2}{2}\omega + \frac{\Lambda^2(V-k)}{(c+\Lambda(V-k))} - \Lambda. \quad (\text{A.14})$$

Next, we prove that $\Pi^{SD} - \Pi^{FIFO}$ is monotone decreasing in ω for $\omega < \tilde{\omega}$. Differentiating $\Pi^{SD} - \Pi^{FIFO}$ with respect to ω gives $\Omega(\omega) \triangleq -\frac{\Lambda^2(V-k)^2}{2(c+\Lambda(V-k))^2\omega^2} + \frac{(\Lambda+c/V)^2}{2}$, which is increasing in ω . Hence, it suffices to show $\Omega(\tilde{\omega}) < 0$. $\Omega(\tilde{\omega}) = -\frac{c^2}{2(V-k)^2} + \frac{(\Lambda+c/V)^2}{2} < 0$, where the last inequality is due to (A.12). Since $\Pi^{SD} - \Pi^{FIFO}$ is monotone decreasing in ω for $\omega < \tilde{\omega}$ and $\Pi^{SD} - \Pi^{FIFO} > 0$ as $\omega \rightarrow 0$ and $\Pi^{SD} - \Pi^{FIFO} < 0$ when $\omega = \tilde{\omega}$, there exists a unique $\hat{\omega}$ such that $\Pi^{SD} - \Pi^{FIFO} > 0$ if and only if $\omega < \hat{\omega}$. Therefore, the optimal mechanism is strategic delay for $\omega < \hat{\omega}$, and non-referral FIFO for $\omega \geq \hat{\omega}$. This proves Case (iii) in the Proposition.

Next, we show that the optimal capacity under strategic delay in the second best, μ^{SD} , is lower than the optimal capacity in the first best, μ^{FB} . Let $\lambda_{FB}(\mu)$ and $\lambda_{SD}(\mu)$ be the throughput as a function of capacity μ in the first best and the second best (with strategic delay), respectively. Thus, μ^{FB} solves $\lambda_{FB}'(\mu^{FB}) - \omega\mu^{FB} = 0$ and μ^{SD} solves $\lambda_{SD}'(\mu^{SD}) - \omega\mu^{SD} = 0$. Note that both $\lambda_{FB}'(\mu)$ and $\lambda_{SD}'(\mu)$ are decreasing in μ . Therefore, to show $\mu^{SD} < \mu^{FB}$, it suffices to show $\lambda_{FB}'(\mu) > \lambda_{SD}'(\mu)$. From (A.9), $\frac{d\mu}{d\lambda_{FB}} = 1 + \frac{c\lambda_{FB}[\lambda_{FB}(V-k)+2k\Lambda]}{\Lambda[\lambda_{FB}(V-k)+k\Lambda]^2}$. From (A.13), $\frac{d\mu}{d\lambda_{SD}} = 1 + \frac{c}{\Lambda(V-k)}$. Since $dy/dx = 1/(dx/dy)$, to show $\lambda_{FB}'(\mu) > \lambda_{SD}'(\mu)$, it suffices to show $\frac{d\mu}{d\lambda_{FB}} < \frac{d\mu}{d\lambda_{SD}}$, i.e., $\frac{c\lambda_{FB}[\lambda_{FB}(V-k)+2k\Lambda]}{\Lambda[\lambda_{FB}(V-k)+k\Lambda]^2} < \frac{c}{\Lambda(V-k)}$. This is equivalent to showing $(V-k)\lambda_{FB}[\lambda_{FB}(V-k)+2k\Lambda] < [\lambda_{FB}(V-k)+k\Lambda]^2$, which is equivalently showing $(k\Lambda)^2 > 0$, which is trivially true. Hence, $\mu^{SD} < \mu^{FB}$.

Case 3: $\bar{h}(0) > 0$ and $\bar{h}(1) > 0$. This gives the condition $\Lambda \in (\underline{\Lambda}, \bar{\Lambda})$. In this case, there exists a unique $r' \in (0, 1)$ such that $\bar{h}(r) > 0$ for $r < r'$; $\bar{h}(r) = 0$ for $r = r'$; and $\bar{h}(r) < 0$ for $r > r'$. Since r^{FB} decreases in ω , it follows that there exists a unique $\underline{\omega}$ such that the first best is achieved by partial priority if $\omega \in (\underline{\omega}, \bar{\omega})$, by full priority if $\omega = \underline{\omega}$, and cannot be achieved in the second best if $\omega < \underline{\omega}$. Next, we show that strategic delay dominates FIFO for $\omega < \underline{\omega}$. First, at $\omega = \underline{\omega}$, $\Pi^{SD}(\underline{\omega}) = \Pi^{FB}(\underline{\omega}) > \Pi^{FIFO}(\underline{\omega})$ and $\mu^{SD}(\underline{\omega}) = \mu^{FB}(\underline{\omega}) > \mu^{FIFO}(\underline{\omega}) = \Lambda + c/V$. Second, from (A.14), $\Pi^{SD}(\omega) - \Pi^{FIFO}(\omega)$ is decreasing in ω if $\omega < \frac{\Lambda(V-k)}{[\Lambda(V-k)+c](\Lambda+c/V)}$. From the expression of μ^{SD} , we have $\mu^{SD}(\underline{\omega}) = \frac{\Lambda(V-k)}{\underline{\omega}[\Lambda(V-k)+c]}$. Since $\mu^{SD}(\underline{\omega}) > \Lambda + c/V$, we have $\underline{\omega} < \frac{\Lambda(V-k)}{[\Lambda(V-k)+c](\Lambda+c/V)}$. Therefore, $\Pi^{SD}(\omega) - \Pi^{FIFO}(\omega)$ is decreasing in ω if $\omega < \underline{\omega}$. Since $\Pi^{SD}(\underline{\omega}) > \Pi^{FIFO}(\underline{\omega})$, we have $\Pi^{SD}(\omega) > \Pi^{FIFO}(\omega)$ for all $\omega < \underline{\omega}$. Thus, strategic delay dominates FIFO and is optimal for $\omega < \underline{\omega}$. This proves Case (ii) of the Proposition. Note that $\mu^{SD} \in (\mu^{FIFO}, \mu^{FB})$ can be similarly shown as before.

Finally, when referrals are generated, the average delay $rW_1 + (1-r)W_0$ is equal to $(V-rk)/c$ because $V-rk - c[rW_1 + (1-r)W_0] = 0$. In a FIFO queue, the average delay W^{FIFO} is equal to V/c . Therefore, $rW_1 + (1-r)W_0 < W^{FIFO}$. This completes the proof. \square

References

Mršević, M (2008) Convexity of the inverse function. *The Teaching of Mathematics* XI(1):21–24.

Appendix B: Equilibrium Conditions for the Full-Priority Referral Scheme

Under the full-priority referral scheme, if a customer makes a successful referral, she gains full priority over all the others who do not make a successful referral. Those with a successful referral (without a successful referral) expect a delay of W_1 (W_0). An equilibrium is specified by customers' joining probability $\alpha \in [0, 1]$ and referral effort $r \in [0, 1]$. In equilibrium, (W_1, W_2) is determined from (α, r) from the following equations:

$$W_1(r, \alpha) = \frac{1}{\mu(1-r\alpha) - \Lambda r \alpha^2}, \quad W_0(r, \alpha) = \frac{\mu(1-r\alpha)}{[\mu(1-r\alpha) - \Lambda r \alpha^2][\mu(1-r\alpha) - \Lambda \alpha]}.$$

Case 1: $(\alpha < 1, r = 0)$ is an equilibrium if $V - \frac{1}{\mu - \Lambda \alpha} = 0$, $c\alpha \left(\frac{1}{\mu - \Lambda \alpha} - \frac{1}{\mu} \right) < k$.

Case 2: $(\alpha = 1, r = 0)$ is an equilibrium if $V - \frac{1}{\mu - \Lambda} \geq 0$, $c \left(\frac{1}{\mu - \Lambda} - \frac{1}{\mu} \right) < k$.

Case 3: $(\alpha < 1, r \in (0, 1))$ is an equilibrium if $V - kr - \frac{1}{\mu(1-r\alpha) - \Lambda \alpha} = 0$, $c\alpha [W_0(r, \alpha) - W_1(r, \alpha)] = k$.

Case 4: $(\alpha = 1, r \in (0, 1))$ is an equilibrium if $V - kr - \frac{1}{\mu(1-r) - \Lambda} \geq 0$, $c[W_0(r, 1) - W_1(r, 1)] = k$.

Case 5: $(\alpha < 1, r = 1)$ is an equilibrium if $V - k - \frac{1}{\mu(1-\alpha) - \Lambda \alpha} = 0$, $c\alpha [W_0(1, \alpha) - W_1(1, \alpha)] \geq k$.

Appendix C: Mechanism Design Formulation that Discriminates Between Base and Referred Customers

The mechanism design problem chooses expected delays $\mathbf{W} = (W_1^B, W_0^B, W_1^R, W_0^R)$ to induce customer strategies $\sigma \triangleq (\alpha^B, \alpha^R, r^B, r^R)$ that maximize the system throughput, subject to the IR, IC, OA and stability constraints. For ease of exposition, let $q = r^B \alpha^R$, $p = r^R \alpha^R$, $\lambda = \Lambda \alpha^B$. The mechanism design formulation is as follows. The objective function maximizes the system throughput: $\max_{\sigma, \mathbf{W}} \frac{\lambda(1+q-p)}{1-p}$. IR constraint 1: $V - kr^B - c[r^B \alpha^R W_1^B + (1 - r^B \alpha^R) W_0^B] \geq 0$. IR constraint 2: $V - kr^R - c[r^R \alpha^R W_1^R + (1 - r^R \alpha^R) W_0^R] \geq 0$. IC constraints: $r^B \in \arg \max_r V - kr' - c[r' \alpha^R W_1^B + (1 - r' \alpha^R) W_0^B]$, $r^R \in \arg \max_r V - kr' - c[r' \alpha^R W_1^R + (1 - r' \alpha^R) W_0^R]$. Stability constraint: $\frac{\lambda(1+q-p)}{1-p} < \mu$. Next, we list fifteen (15) OA constraints. We invoke Lemma B.1 of Yang and Debo (2019): In a batch-arrival queue with exponential service rate μ , Poisson arrival rate λ , and random batch size N with finite first moment $\mathbb{E}[N]$ and second moment $\mathbb{E}[N^2]$, the expected delay W in the system is $W = \frac{\mathbb{E}[N] + \mathbb{E}[N^2]}{2\mathbb{E}[N](\mu - \lambda \mathbb{E}[N])}$. OA constraint 1: $W_1^B \geq \frac{1}{\mu - \lambda q}$. OA constraint 2: $W_0^B \geq \frac{1}{\mu - \lambda(1-q)}$. OA constraint 3: $W_1^R \geq \frac{1}{\mu(1-p) - \lambda qp}$. OA constraint 4: $W_0^R \geq \frac{1}{\mu - \lambda q}$. OA constraint 5: $qW_1^B + (1-q)W_0^B \geq \frac{1}{\mu - \lambda}$. OA constraint 6:

$$\lambda q W_1^B + \frac{\lambda q p}{1-p} W_1^R \geq \frac{\lambda q}{1-p} \frac{1}{\mu(1-p) - \lambda q}. \quad (\text{C.1})$$

OA constraint 7: $\lambda q W_1^B + \lambda q W_0^R \geq 2\lambda q \frac{3}{2[\mu - 2\lambda q]}$. OA constraint 8: $\frac{\lambda q p}{1-p} W_1^R + \lambda(1-q)W_0^B \geq \frac{\lambda(1-p-q+2pq)}{1-p} \frac{pq+(1-q)(1-p)^2}{(1-p-q+2pq)(\mu(1-p) - \lambda(1-p-q+2pq))}$. OA constraint 9: $\frac{\lambda q p}{1-p} W_1^R + \lambda q W_0^R \geq \frac{\lambda q}{1-p} \frac{1}{\mu(1-p) - \lambda q}$. OA constraint 10: $\lambda(1-q)W_0^B + \lambda q W_0^R \geq \frac{\lambda}{\mu - \lambda}$. OA constraint 11: $\lambda q W_1^B + \frac{\lambda q p}{1-p} W_1^R + \lambda(1-q)W_0^B \geq \frac{\lambda(1-p+pq)}{1-p} \frac{(1-q)(1-p)^2+q}{(1-p+pq)[\mu(1-p) - \lambda(1-p+pq)]}$. OA constraint 12: $\lambda q W_1^B + \frac{\lambda q p}{1-p} W_1^R + \lambda q W_0^R \geq \frac{\lambda q(2-p)}{1-p} \frac{(1-p)^2+2}{2(2-p)[\mu(1-p) - \lambda(2-p)]}$. OA constraint 13: $\frac{\lambda q p}{1-p} W_1^R + \lambda(1-q)W_0^B + \lambda q W_0^R \geq \frac{\lambda(1-p+pq)}{1-p} \frac{(1-q)(1-p)^2+q}{(1-p+pq)[\mu(1-p) - \lambda(1-p+pq)]}$. OA constraint 14: $\lambda q W_1^B + \lambda(1-q)W_0^B + \lambda q W_0^R \geq \lambda(1+q) \frac{1+2q}{(1+q)(\mu - \lambda(1+q))}$. OA constraint 15:

$$\lambda q W_1^B + \frac{\lambda q p}{1-p} W_1^R + \lambda(1-q)W_0^B + \lambda q W_0^R \geq \frac{\lambda(1+q-p)}{1-p} \frac{1+q-p+(1-p)(q-p)}{(1+q-p)[\mu(1-p) - \lambda(1+q-p)]}. \quad (\text{C.2})$$

In terms of the structure of the optimal referral mechanism, partial priority (subject to work conservation) corresponds to (C.1) being non-binding and (C.2) being binding; strategic delay (with full priority assigned to referring customers) corresponds to (C.1) being binding and (C.2) being non-binding; full priority (subject to work conservation) corresponds to both (C.1) and (C.2) being binding; non-referral FIFO corresponds to $r^B = r^R = 0$.