Online Appendices

A Proof of Lemma 1.

Consider an arbitrary facility $i \in \{1, 2, ..., N\}$, and let $g_i(T, W)$ denote the expected long-run average reward under a threshold policy with threshold $T \in \mathbb{N}_0$ and admission charge $W \in \mathbb{R}$. Then:

$$g_i(T, W) = \lambda(1 - \pi_i(T, T))(\alpha_i - W) - \beta_i \sum_{y=0}^{T} y\pi_i(y, T). \tag{34}$$

It follows that, given any fixed $T \in \mathbb{N}_0$, the average reward $g_i(T, W)$ is a linear, strictly decreasing function of W, and its gradient is:

$$\frac{\partial g_i}{\partial W} = -\lambda(1 - \pi_i(T, T)). \tag{35}$$

Given some admission charge $W \in \mathbb{R}$, let $T_i^*(W)$ be the optimal threshold referred to in Definition 1; hence, $g_i(T_i^*(W), W) \geq g_i(T, W)$ for all $T \in \mathbb{N}_0$. It can be verified using standard formulae for finite-capacity $M/M/c$ queues (see Gross and Harris (1998), p. 74) that the steady-state probability $\pi_i(T, T)$ is strictly decreasing with T; hence, the gradient in (35) is also strictly decreasing with T (see Figure 3). Given that $g_i(T_i^*(W), W) \geq g_i(T_i^*(W) + n, W)$ for arbitrary $n \geq 1$, it must therefore be the case that $g_i(T_i^*(W), W') > g_i(T_i^*(W) + n, W')$ for any $W' > W$, and therefore the policy with threshold $T + n$ cannot be optimal under an admission charge W', since it does not maximize the average reward. It follows that $T_i^*(W') \leq T_i^*(W)$ for any two admission charges $W, W' \in \mathbb{R}$ with $W < W'$, which verifies the first of the indexability conditions.

Using similar arguments, for any state $x \in \mathbb{N}$ there must exist some value $W_i(x) \in \mathbb{R}$ such that $g_i(x, W) \geq g_i(x - n, W)$ for all $n \in \{1, 2, ..., x\}$ if and only if $W \leq W_i(x)$. This is due to the fact that the linear functions $g_i(x - n, W)$ have larger gradients than that of $g_i(x, W)$. Hence, it must be the case that $T_i^*(W) > x$ if and only if $W < W_i(x)$, since this is the only scenario in which a threshold policy with threshold greater than x performs better than an x-threshold policy. This completes the proof that the conditions stated in Definition 1 are both satisfied.

In order to derive an expression for $W_i(x)$, we note that $W_i(x)$ must be the unique value of W which results in the thresholds $T = x$ and $T = x + 1$ both yielding the same expected long-run average reward in a single-facility problem. By equating long-run average rewards under the thresholds x and

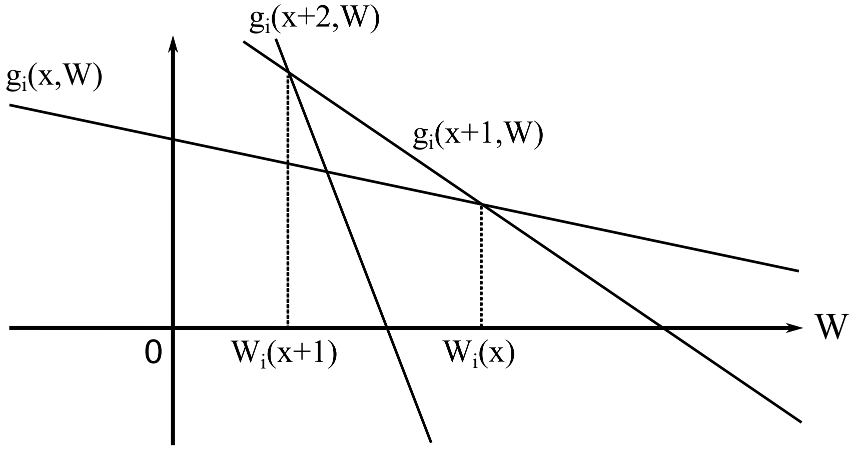

Fig. 3: The linear dependence of the functions $g_i(x, W)$ on the admission charge W.

$x + 1$, we find that $W_i(x)$ satisfies the equation

$$\lambda(1 - \pi_i(x, x))(\alpha_i - W_i(x)) - \beta_i \sum_{y=0}^{x} y\pi_i(y, x)$$
$$= \lambda(1 - \pi_i(x + 1, x + 1))(\alpha_i - W_i(x)) - \beta_i \sum_{y=0}^{x+1} y\pi_i(y, x + 1),$$

Solving this equation directly for $W_i(x)$ yields (13). □

B Proof of the lower bound $S^\circ \subseteq S_{\theta^*}$.

Here we prove the claim that any optimal stationary policy θ^* satisfies the relationship $S^\circ \subseteq S_{\theta^*}$, where S° is defined in (16) as the set of states with no customers waiting in queues.

We will begin by showing that an optimal stationary policy θ^* cannot choose to balk at any state $\mathbf{x} \in S_{\theta^*}$ with $x_i < c_i$ for some $i \in \{1, 2, ..., N\}$. Since the state $\mathbf{0}$ (with no customers present at any facilities) is positive recurrent under all stationary policies, there must be some point in time at which the system finds itself in a state $\mathbf{x} \in S_{\theta^*}$ with $x_i < c_i$ for at least one $i \in \{1, 2, ..., N\}$ while operating under θ^*. In the first part of this proof, $\mathbf{x}$ will denote a fixed state in S_{θ^*} with $x_i < c_i$ for a particular facility i.

We will use a sample path argument. Let ψ denote a non-stationary policy which operates by 'copying' the actions of policy θ^* at all times, *unless* either of the following scenarios apply:

1. θ^* chooses to balk at the state $\mathbf{x}$ (with $x_i < c_i$), in which case ψ chooses to join facility i if the process following policy ψ is also in state $\mathbf{x}$, and otherwise chooses to balk;
2. θ^* chooses to join facility i at some point in time at which the process following policy ψ has more customers at i than the process following θ^*, in which case ψ chooses to balk.

Let $(\mathbf{x}_n)_{n\in\mathbb{N}_0}$ and $(\mathbf{y}_n)_{n\in\mathbb{N}_0}$ denote the state evolutions of two processes Υ_1 and Υ_2 which follow policies θ^* and ψ respectively. Then, in notation, ψ operates as follows:

$$\psi(\mathbf{y}_n) = \begin{cases} i, & \text{if } \mathbf{x}_n = \mathbf{y}_n = \mathbf{x}, \\ 0, & \text{if } \theta^*(\mathbf{x}_n) = i \text{ and } (\mathbf{x}_n)_i < (\mathbf{y}_n)_i, \\ \theta^*(\mathbf{x}_n), & \text{otherwise,} \end{cases}$$

where $(\mathbf{x}_n)_i$ and $(\mathbf{y}_n)_i$ denote the i^{th} components of $\mathbf{x}_n$ and $\mathbf{y}_n$ respectively. Suppose that both of the processes Υ_1 and Υ_2 are initialized in the regenerative state $\mathbf{0}$ and evolve according to the same sequence of random events. Let $n_0 \geq 0$ denote the first discrete time step at which both processes are in state $\mathbf{x}$ and an arrival occurs. Then $n_0 + 1$ is the first step at which the two processes differ. Noting that $x_i < c_i$, it follows from the reward formulation (2) that the process Υ_2 following ψ earns a strictly greater single-state reward at step $n_0 + 1$ than the process Υ_1, since it has an extra customer being served at facility i. Furthermore, let n_1 denote the next time epoch (after n_0) at which the processes Υ_1 and Υ_2 are once again in the same state; this may occur as a result of a service completion at facility i being 'seen' by Υ_2 but not by Υ_1, or as a result of θ^* choosing to join facility i while the process Υ_2 already has an extra customer present at i. In either case, it can easily be checked using the definition of policy ψ that

$$(\mathbf{x}_n)_i < (\mathbf{y}_n)_i \leq c_i \quad \forall n \in [n_0 + 1, n_1 - 1].$$

Noting that $n_1 \geq n_0 + 2$ (since the two processes must first diverge from each other before meeting again), it follows from (2) that Υ_2 earns a strictly greater total reward than Υ_1 over the interval $[n_0, n_1]$. Since the state $\mathbf{x}$ (being accessible from $\mathbf{0}$) is positive recurrent under θ^*, the two processes make infinitely many visits to $\mathbf{x}$ with probability one. It follows that the policy ψ is superior to θ^* with respect to the average reward criterion, which contradicts the fact that θ^* is an optimal policy.

The sample path argument given above proves that if $\mathbf{x} \in S_{\theta^*}$ is a state with $x_i < c_i$ for some facility i, then the optimal stationary policy θ^* cannot choose to balk at $\mathbf{x}$. To complete the proof, we must show that all states with $x_i \leq c_i$ for all $i \in \{1, 2, ..., N\}$ are included in S_{θ^*}. Indeed, suppose the system is in the regenerative state $\mathbf{0}$ and a sequence of $\tilde{B} + 1$ consecutive customer arrivals occurs (without any service completions), where $\tilde{B} = \sum_{i=1}^{N} \tilde{B}_i$ and $\tilde{B}_i = \lfloor \beta_i / (\alpha_i c_i \mu_i) \rfloor$. Since it is proved in Shone et al. (2016) that S_{θ^*} is contained in the selfish state space $\tilde{S}$, it follows from the definition of $\tilde{S}$ in (5) that θ^* must choose to balk at least once during the sequence of $\tilde{B}+1$ arrivals, otherwise the process would pass outside $\tilde{S}$. Due to our previous sample path argument, we know that θ^* will not choose to balk at a state which has any servers idle, so therefore it must balk at some state $\mathbf{x} \in S_{\theta^*}$ which has $x_i \geq c_i$ for all facilities i. The state $\mathbf{x}$ in question is positive recurrent under θ^* (since it is accessible from $\mathbf{0}$), and therefore all states $\mathbf{y}$ which satisfy the componentwise inequality $y_i \leq x_i$ for all $i \in \{1, 2, ..., N\}$ are also positive recurrent under θ^* since they are accessible from $\mathbf{x}$ via service completions. This completes the proof. □

C Proof of Lemma 2.

The proof is accomplished using dynamic programming (DP) arguments. Let θ^* be an optimal stationary policy. It is useful to note that, due to Theorem 2 in (Shone et al. (2016)), θ^* can be obtained as an optimal solution to a *finite*-state problem with state space $\tilde{S}$ defined in (5). In this finite-state problem we assume that the set of actions $A_{\mathbf{x}}$ available at state $\mathbf{x} \in \tilde{S}$ excludes any facilities $i \in \{1, 2, ..., N\}$ for which $x_i = \lfloor \beta_i / (\alpha_i c_i \mu_i) \rfloor$. Results for finite-state MDPs (see Puterman (1994)) imply that there exist a constant g^* and a real-valued function h satisfying the average reward optimality equations

$$g^* + h(\mathbf{x}) = \max_{a \in A_{\mathbf{x}}} \left\{ r(\mathbf{x}) + \sum_{\mathbf{y} \in \tilde{S}} p(\mathbf{x}, a, \mathbf{y}) h(\mathbf{y}) \right\} \quad (\mathbf{x} \in \tilde{S}). \tag{36}$$

In these equations, g^* is the optimal average reward and h is the relative value function, which is unique up to an additive constant. For states $\mathbf{x}$ in the recurrent set S_{θ^*} it must be the case that θ^* chooses an action which attains the maximum on the right-hand side of (36); otherwise, one would be able to obtain a policy superior to θ^* using policy iteration.

In this proof we consider two different problems simultaneously. The first is the N-facility problem formulated in Section 2, and the second is a single-facility problem involving only facility $i \in \{1, 2, ..., N\}$, where (throughout this proof) we regard the facility i as fixed but arbitrary. For the single-facility problem we will write the optimality equations as

$$G^* + H(x) = \max_{a \in A_x} \left\{ R(x) + \sum_{y \in \tilde{S}_i} P(x, a, y) H(y) \right\} \quad (y \in \tilde{S}_i), \tag{37}$$

where G^*, $H(\cdot)$, $R(\cdot)$ and $P(\cdot)$ are the analogues of g^*, $h(\cdot)$, $r(\cdot)$ and $p(\cdot)$ pertaining to a single-facility problem with facility i and $\tilde{S}_i = \{0, 1, ..., \tilde{T}_i\}$, where $\tilde{T}_i = \lfloor \beta_i / (\alpha_i c_i \mu_i) \rfloor \}$. We will use induction based on dynamic programming (DP) value iteration and consider the functions h and H obtained as limits of the finite-stage iterates h_k and H_k as $k \to \infty$. We will suppose that, in both problems, the discrete-time step size is $\Delta := (\lambda + \sum_{j=1}^{N} c_j \mu_j)^{-1}$. (This step size can be used in the single-facility problem because it is smaller than the maximum step size of $(\lambda + c_i\mu_i)^{-1}$.) Our aim is to show that, given an arbitrary state $\mathbf{x} \in \tilde{S}$, the following holds for all integers y satisfying $x_i \le y < \tilde{T}_i$ (where x_i is the i^{th} component of $\mathbf{x}$):

$$h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) \ge H_k(y+1) - H_k(y) \quad \forall k \in \mathbb{N}_0. \tag{38}$$

We note that (38) holds trivially when $k = 0$ because $h_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \tilde{S}$ and $H_k(x) = 0$ for all $x \le \tilde{T}_i$. We assume that it holds for arbitrary $k \in \mathbb{N}_0$ and aim to show that it holds with k replaced by $k+1$. By enumerating transition probabilities, we obtain the following:

$$\begin{aligned}
&h_{k+1}(\mathbf{x}^{i+}) - h_{k+1}(\mathbf{x}) - H_{k+1}(y+1) + H_{k+1}(y) \\
&= r(\mathbf{x}^{i+}) - r(\mathbf{x}) - R(y+1) + R(y) && (39)\\
&\quad + \lambda\Delta\left(h_k((\mathbf{x}^{i+})^{a_1+}) - h_k(\mathbf{x}^{a_0+}) - H_k(y+1+b_1) + H_k(y+b_0)\right) && (40)\\
&\quad + \sum_{j\ne i}\min(x_j,c_j)\mu_j\Delta\left(h_k((\mathbf{x}^{i+})^{j-}) - h_k(\mathbf{x}^{j-}) - H_k(y+1) + H_k(y)\right) && (41)\\
&\quad + \min(x_i,c_i)\mu_i\Delta\left(h_k(\mathbf{x}) - h_k(\mathbf{x}^{i-}) - H_k(y) + H_k(y-1)\right) && (42)\\
&\quad + I(x_i < c_i \text{ and } x_i < y)\mu_i\Delta\left(h_k(\mathbf{x}) - h_k(\mathbf{x}) - H_k(y) + H_k(y-1)\right) && (43)\\
&\quad + I(x_i = y < c_i)\mu_i\Delta\left(h_k(\mathbf{x}) - h_k(\mathbf{x}) - H_k(y) + H_k(y)\right) && (44)\\
&\quad + I(x_i < y-1 \text{ and } x_i < c_i - 1)(\min(y,c_i) - (x_i+1))\mu_i\Delta \\
&\qquad \times \left(h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) - H_k(y) + H_k(y-1)\right) && (45)\\
&\quad + I(x_i < y < c_i)\mu_i\Delta\left(h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) - H_k(y) + H_k(y)\right) && (46)\\
&\quad + \left(\max(c_i - (y+1), 0)\mu_i\Delta + \sum_{j\ne i}\max(c_j - x_j, 0)\mu_j\Delta\right) \\
&\qquad \times \left(h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) - H_k(y+1) + H_k(y)\right), && (47)
\end{aligned}$$

where $a_0, a_1 \in \{0, 1, ..., N\}$ are optimal actions at states $\mathbf{x}$ and $\mathbf{x}^{i+}$ in a $(k+1)$-stage problem with N facilities, and similarly $b_0, b_1 \in \{0, 1\}$ are optimal actions at states y and $y+1$ in the single-facility problem with $k+1$ stages. (We adopt the notational convention that $\mathbf{x}^{0+} = \mathbf{x}$ for $\mathbf{x} \in S$.)

In order to make sense of the above equation, the reader may wish to think of a sample path argument, in which the proof essentially works in an analogous way. One may consider 4 random processes, denoted by Υ_1, Υ_2, Υ_3 and Υ_4. Processes Υ_1 and Υ_2 are initialized at states $\mathbf{x}^{i+}$ and $\mathbf{x}$ respectively in a system with N facilities, while Υ_3 and Υ_4 are initialized at states $y+1$ and y respectively in a single-facility system. All four processes evolve according to the same sequence of random events and follow optimal decision-making policies in a finite-stage problem with $k+1$ stages. Line (40) represents all four processes receiving an arrival, while line (41) represents service completions at facilities $j \ne i$ 'seen' by Υ_1 and Υ_2, but not seen by Υ_3 or Υ_4 (because the facilities in question do not exist in the single-facility system). Line (42) represents all four processes seeing a service completion at facility i (which, in the case of Υ_3 and Υ_4, is the only facility). Line (43) represents a service completion seen by all processes except Υ_2, which is possible if $x_i < c_i$ and $x_i < y$ (in which case Υ_2 has the smallest number of services in progress at i). Line (44) represents a service completion at i seen by Υ_1 and Υ_3, but not by Υ_2 or Υ_4 (which is possible only if $x_i = y < c_i$, so that Υ_1 and Υ_3 jointly have the greatest number of services in progress at i). Line (45) represents service completions at i seen by Υ_3 and Υ_4 but not by Υ_1 or Υ_2, which requires x_i+1 to be smaller than both y and c_i. Line (46) represents a service completion seen by Υ_3 only, which requires $x_i < y < c_i$ (so that Υ_3 has the greatest number of services in progress at i). Line (47) represents service completions seen by none of the four processes.

The objective is to show that the sum of the terms in lines (39)-(47) is non-negative. Using (2), the reward terms in line (39) reduce to:

$$r(\mathbf{x}^{i+}) - r(\mathbf{x}) - R(y+1) + R(y) = \begin{cases} \alpha_i \mu_i, & \text{if } x_i < c_i \leq y, \\ 0, & \text{otherwise.} \end{cases} \tag{48}$$

Hence, line (39) is non-negative. Since $\mathbf{x}$ is a state whose i^{th} component is bounded above by y, the same also applies to the state $\mathbf{x}^{j-}$ (where $j \in \{1, 2, ..., N\} \setminus \{i\}$), and hence line (41) is non-negative due to the inductive assumption (38). Using similar reasoning, $\mathbf{x}^{i-}$ is a state whose i^{th} component is bounded above by $y-1$ (due to the inequality $x_i \leq y$) and therefore line (42) is non-negative by the inductive assumption. Line (44) is trivially equal to zero. The indicator term $I(x_i < y-1 \text{ and } x_i < c_i - 1)$ can only be non-zero if $x_i < y$, in which case the inequality $x_i \leq y-1$ holds, implying that line (45) is also non-negative by the inductive assumption. Line (47) is obviously also non-negative, again due to the inductive assumption.

Summarizing the arguments given thus far, in order to complete the inductive proof that (38) holds with k replaced by $k = 1$ it remains only to show that lines (40), (43) and (46) cannot cause the sum of the terms in lines (39)-(47) to be negative. Consider line (46) first. The indicator term $I(x_i < y < c_i)$ can only be non-zero if the indicator $I(x_i < c_i \text{ and } x_i < y)$ in line (43) is also non-zero, in which case the sum of lines (46) and (43) is equal to $\mu_i \Delta \left(h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) - H_k(y) + H_k(y-1) \right)$, which is again non-negative by the inductive assumption due to the fact that (in this particular case) the inequality $x_i \leq y-1$ holds.

Next, consider line (43) and assume $x_i < c_i$ and $x_i < y$ to avoid triviality. If $y < c_i$ then, as explained in the paragraph above, the sum of lines (43) and (46) is non-negative. On the other hand, suppose $y \geq c_i$. In this case (48) implies that the reward terms in line (39) yield a positive term $\alpha_i \mu_i$. Hence, summing lines (43) and (39) gives:

$$\alpha_i \mu_i + \mu_i \Delta \left(H_k(y-1) - H_k(y) \right). \tag{49}$$

However, the upper boundedness property (72) shown in the proof of Theorem 4 (which applies to any single-facility system) implies that the difference $H_k(y) - H_k(y-1)$ satisfies:

$$H_k(y) - H_k(y-1) \leq \frac{\alpha_i \mu_i - \beta_i}{\mu_i \Delta},$$

which immediately implies that the expression in (49) is bounded *below* by the positive quantity β_i.

These arguments show that the sum of the terms in lines (39), (43) and (46) is always non-negative. In order to complete the proof that (38) is valid, it remains only to show that the 'arrival terms' in line (40) sum to a non-negative value. This may be done by considering all possibilities for the actions a_0 and b_1 and showing, in each case, that there exists some choice for the actions a_1 and b_0 such that line (40) is non-negative (this approach is valid because, in practice, a_1 and b_0 are *maximizing* actions). These possibilities are considered below.

- If $a_0 = b_1 = 0$, we consider $a_1 = b_0 = 0$, in which case line (40) is non-negative by the inductive assumption.
- Similarly, if $a_0 = i$ and $b_1 = 1$ then we consider $a_1 = i$ and $b_0 = 1$, in which case line (40) is again non-negative by the inductive assumption since $x_i + 1 \leq y + 1$.
- If $a_0 = i$ and $b_1 = 0$ then we consider $a_1 = 0$ and $b_0 = 1$, in which case line (40) is trivially equal to zero.
- If $a_0 = j$ for some $j \in \{1, 2, ..., N\} \setminus \{i\}$ and $b_1 = 0$ then we consider $a_1 = j$ and $b_0 = 0$, in which case line (40) is again non-negative by the inductive assumption since $\mathbf{x}^{j+}$ is a state whose i^{th} component is bounded above by y.
- If $a_0 = j$ for some $j \in \{1, 2, ..., N\} \setminus \{i\}$ and $b_1 = 1$ then we consider $a_1 = j$ and $b_0 = 1$, with the result that line (40) is non-negative by the inductive assumption since $\mathbf{x}^{j+}$ is a state whose i^{th} component is bounded above by $y + 1$.
- Finally, if $a_0 = 0$ and $b_1 = 1$, then this implies $h_k(\mathbf{x}^{i+}) \leq h_k(\mathbf{x})$ and $H_k(y+2) \geq H_k(y+1)$. The latter inequality implies $H_k(y+1) \geq H_k(y)$ due to the concavity property (73) shown in the proof of Theorem 4 (which applies to any single-facility system). However, due to the inductive assumption (38), we then have $H_k(y+2) - H_k(y+1) = h_k(\mathbf{x}^{i+}) - h_k(\mathbf{x}) = 0$. Therefore, by considering $a_1 = 0$ and $b_0 = 1$, we find that line (40) is bounded below by zero.

Thus, it has been verified that line (40) is non-negative in all possible cases. In view of the previous arguments, this establishes that the sum of the terms in lines (39)-(47) must be non-negative, which completes the inductive proof that (38) holds for all $k \in \mathbb{N}_0$. By the limiting properties of $h_n(\cdot)$ and $H_n(\cdot)$ as $n \to \infty$, this implies that

$$h(\mathbf{x}^{i+}) - h(\mathbf{x}) \geq H(y+1) - H(y) \tag{50}$$

for all $\mathbf{x} \in \tilde{S}$ and $y \in \mathbb{N}_0$ with $x_i \leq y < \tilde{T}_i$, where $i \in \{1, 2, ..., N\}$ is arbitrary. We now complete the proof by appealing to the decision-making properties of θ^* and $\theta^{[W]}$. Given that θ^* balks at the recurrent state $\mathbf{x} \in S_{\theta^*}$, the optimality equations (36) imply $h(\mathbf{x}) \leq h(\mathbf{x}^{i+})$ for any facility $i \in \{1, 2, ..., N\}$ such that $\mathbf{x}^{i+} \in \tilde{S}$. Due to (50), we therefore have $H(y) \leq H(y+1)$ for all $y \in \{x_i, x_i{+}1, ..., \tilde{T}_i{-}1\}$ and therefore the single-facility optimality equations imply the existence of an optimal policy which balks under these states. By definition of the Whittle indices, it follows that $W_i(y) \leq 0$ for $y \in \{x_i, x_i+1, ..., \tilde{T}_i - 1\}$, and so the Whittle policy $\theta^{[W]}$ declines to join facility i under state $\mathbf{x}$. Since the facility i was arbitrary, we conclude that $\theta^{[W]}$ balks at state $\mathbf{x}$. This completes the proof. □

D Proof of Theorem 2.

In this proof we will use $W_i(x, \lambda)$ to denote the Whittle index $W_i(x)$ (defined in (13)) given a demand rate λ. We begin by considering the light-traffic scenario. Let $\mathbf{0}$ denote the state in S with no customers present at any facility. By setting $x = 0$ in (13) and using standard formulae for the probabilities $\pi_i(y, x)$ one can verify that

$$W_i(0, \lambda) = \alpha_i - \frac{\beta_i}{\mu_i}. \tag{51}$$

This states, logically, that a 0-threshold policy is preferred to a 1-threshold policy at facility i if and only if the admission charge is greater than a customer's expected net reward for joining when a server is available. It then follows from (51) (and the assumption that $\alpha_i - \beta_i/\mu_i > 0$ for each facility i) that, regardless of the demand rate λ, the Whittle policy $\theta^{[W]}(\lambda)$ always chooses a facility at state $\mathbf{0}$ which maximizes $\alpha_i - \beta_i/\mu_i$.

Recall that we are considering a uniformized system with step size $(\lambda + \sum_{i=1}^{N} c_i\mu_i)^{-1} = 1$, and therefore λ can be interpreted as the probability that a customer arrives in any particular discrete time step. This implies that we can express the long-run average reward g^θ under an arbitrary stationary policy θ as

$$g^\theta = \lambda \sum_{\mathbf{x} \in S} \pi_\theta(\mathbf{x}) w_{\theta(\mathbf{x})}(\mathbf{x}), \tag{52}$$

where $\pi_\theta(\mathbf{x})$ is the stationary probability for state $\mathbf{x}$ under θ and $w_a(\mathbf{x})$ is a customer's expected net reward for choosing action $a \in \{0, 1, ..., N\}$ under state $\mathbf{x}$, defined in (4). Indeed, it is shown rigorously in (Shone et al. (2016)) (Lemma 1) that one can replace the reward function (2) with an alternative, action-dependent function $\hat{r}(\mathbf{x}, a) = \lambda w_a(\mathbf{x})$ in the MDP formulation for the problem without affecting the performance measure g^θ under any stationary policy θ. Using (52), we can write

$$\frac{g^*(\lambda) - g^{[W]}(\lambda)}{g^*(\lambda)} = \frac{\sum_{\mathbf{x} \in S} \pi_{\theta^*}(\mathbf{x}) w_{\theta^*(\mathbf{x})}(\mathbf{x}) - \sum_{\mathbf{x} \in S} \pi_{\theta^{[W]}}(\mathbf{x}) w_{\theta^{[W]}(\mathbf{x})}(\mathbf{x})}{\sum_{\mathbf{x} \in S} \pi_{\theta^*}(\mathbf{x}) w_{\theta^*(\mathbf{x})}(\mathbf{x})}, \tag{53}$$

where θ^* denotes an optimal stationary policy. Trivially, one can show that $\lim_{\lambda \to 0} \pi_\theta(\mathbf{0}) = 1$ for any $\theta \in \Theta$ and hence the quantity on the right-hand side of (53) tends to

$$\frac{w_{\theta^*}(\mathbf{0}) - w_{\theta[W]}(\mathbf{0})}{w_{\theta^*}(\mathbf{0})} \tag{54}$$

as $\lambda \to 0$. We note that, for any $\lambda > 0$,

$$\frac{g^{[W]}(\lambda)}{\lambda} \leq \frac{g^*(\lambda)}{\lambda} \leq \max_i \left\{ \alpha_i - \frac{\beta_i}{\mu_i} \right\},$$

where the first inequality is by definition of an optimal policy and the second is due to the fact that no policy can earn an average reward greater than $\lambda \max_i \{\alpha_i - \beta_i \mu_i\}$, as shown by (52). By taking limits as $\lambda \to 0$, we obtain

$$w_{\theta[W]}(\mathbf{0}) \leq w_{\theta^*}(\mathbf{0}) \leq \max_i \left\{ \alpha_i - \frac{\beta_i}{\mu_i} \right\}. \tag{55}$$

By the previous arguments, $w_{\theta[W]}(\mathbf{0})$ is equal to $\max_i \{\alpha_i - \beta_i/\mu_i\}$, so it follows from (55) that $w_{\theta^*}(\mathbf{0}) - w_{\theta[W]}(\mathbf{0})$ tends to zero as $\lambda \to 0$. In view of (54), this establishes the result.

In the heavy-traffic case, we recall that the Whittle policy is based on an optimal policy for a relaxation of the problem, which involves applying optimal, independent threshold policies at the N facilities individually. By analogy to (2), we can say that the reward function for the single-facility problem represented in Figure 2 is

$$r_i(x) = (\alpha_i - W)\min(x, c_i)\mu_i - \beta_i x. \tag{56}$$

Under an n-threshold policy, it is trivial to show that the stationary probability for state n tends to 1 as $\lambda \to \infty$. Suppose we apply an x-threshold policy at facility i, where $x < c_i$. As $\lambda \to \infty$, we approach a situation where the system is always in state x and therefore, by (56), the long-run average reward under this policy approaches

$$x[(\alpha_i - W)\mu_i - \beta_i].$$

By the arguments in Section 3, the Whittle index $W_i(x, \lambda)$ (in the heavy-traffic limit) is obtained by solving the equation

$$x[(\alpha_i - W)\mu_i - \beta_i] = (x+1)[(\alpha_i - W)\mu_i - \beta_i]$$

for W and hence we obtain

$$\lim_{\lambda \to \infty} W_i(x, \lambda) = \alpha_i - \beta_i/\mu_i,$$

which is positive by the assumptions of our model. This states that, in the heavy-traffic limit, the Whittle policy (with respect to facility i) always directs a customer to join rather than balk if there is an idle server available. On the other hand, if we consider a state $y \geq c_i$, then it can easily be seen from (56) that $r_i(y)$ is strictly greater than $r_i(y+1)$, and therefore (in the heavy-traffic limit) a y-threshold policy will always earn a strictly greater average reward than a $(y+1)$-threshold policy, regardless of the value of W. This means that the Whittle index $W_i(y, \lambda)$ tends to minus infinity, since balking is preferred to joining at state y for all $y \geq c_i$.

Since these arguments apply to all facilities $i \in \{1, 2, ..., N\}$ it follows that, as $\lambda \to \infty$, we obtain a Whittle heuristic policy which directs a customer to balk if and only if $x_i \geq c_i$ for all $i \in \{1, 2, ..., N\}$ (i.e. all servers are busy at all facilities). As $\lambda \to \infty$, the average reward under this policy tends to

$$\sum_{i=1}^{N} c_i(\alpha_i \mu_i - \beta_i). \tag{57}$$

However, examining the reward function (2) in our MDP formulation for the problem, we see that it is uniformly bounded above by the expression in (57) and therefore (57) also represents the limiting value for the optimal average reward as $\lambda \to \infty$. Thus, we have established that the Whittle heuristic attains optimality in a heavy-traffic limit. □

E Proof of Theorem 3 (Statement 2).

We will use an inductive proof, based on the average reward optimality equations (19), to show that the function h found by value iteration possesses the properties of concavity, submodularity and diagonal submissiveness in the special case of two facilities with a single server at each. The inductive method is based on the relative value functions $h_n(\mathbf{x}) = v_n(\mathbf{x}) - v_n(\mathbf{0})$ (for $\mathbf{x} \in S$ and $n \in \mathbb{N}_0$), where v_n satisfies the finite-horizon optimality equations:

$$v_{n+1}(\mathbf{x}) = \max_{a \in A} \left\{ r(\mathbf{x}) + \sum_{\mathbf{y} \in S} p(\mathbf{x}, a, \mathbf{y}) v_n(\mathbf{y}) \right\} \quad (\mathbf{x} \in S). \tag{58}$$

For ease of notation, let us define the first-order differences $D_j(\mathbf{x}, f)$ for real-valued functions f, facilities $j \in \{1, 2\}$ and states $\mathbf{x} \in S$ as follows:

$$D_j(\mathbf{x}, f) := f(\mathbf{x}^{j+}) - f(\mathbf{x}).$$

We will also define the *second-order* differences $D_{jj}(\mathbf{x}, f)$ and $D_{ij}(\mathbf{x}, f)$ as follows:

$$D_{jj}(\mathbf{x}, f) := D_j(\mathbf{x}, D_j(\mathbf{x}, f)) = f((\mathbf{x}^{j+})^{j+}) - f(\mathbf{x}^{j+}) - f(\mathbf{x}^{j+}) + f(\mathbf{x}),$$
$$D_{ij}(\mathbf{x}, f) := D_i(\mathbf{x}, D_j(\mathbf{x}, f)) = f((\mathbf{x}^{i+})^{j+}) - f(\mathbf{x}^{i+}) - f(\mathbf{x}^{j+}) + f(\mathbf{x}).$$

It is sufficient to show that for states $\mathbf{x} \in S$, $i, j \in \{1, 2\}$ with $i \neq j$ and $n \in \mathbb{N}_0$:

$$D_{jj}(\mathbf{x}, h_n) \leq 0, \tag{59}$$
$$D_{ij}(\mathbf{x}, h_n) \leq 0, \tag{60}$$
$$D_{jj}(\mathbf{x}, h_n) - D_{ij}(\mathbf{x}, h_n) \leq 0. \tag{61}$$

We define $v_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in S$, and hence all three properties hold trivially when $n = 0$. Assume that they also hold when $n = k$, where $k \in \mathbb{N}_0$ is arbitrary. The submodularity property (60) will be considered first. At stage $n = k + 1$, one has (after simplifications):

$$\begin{aligned} D_{ij}(\mathbf{x}, h_{k+1}) = D_{ij}(\mathbf{x}, r) &+ \lambda\Delta \left(h_k(((\mathbf{x}^{i+})^{j+})^{a_1+}) - h_k((\mathbf{x}^{i+})^{a_0+}) - h_k((\mathbf{x}^{j+})^{b_1+}) + h_k(\mathbf{x}^{b_0+})\right) \\ &+ \sum_{m=1}^{2} \min(x_m, 1)\mu_m \Delta D_{ij}(\mathbf{x}^{m-}, h_k) \\ &+ \left(1 - \lambda\Delta - \sum_{m=1}^{2} \min(x_m, 1)\mu_m\Delta - I(x_i = 0)\mu_i\Delta - I(x_j = 0)\mu_j\Delta\right) D_{ij}(\mathbf{x}, h_k), \end{aligned}$$

where a_0, a_1, b_0, b_1 are optimal actions at states $\mathbf{x}^{i+}$, $(\mathbf{x}^{i+})^{j+}$, $\mathbf{x}$ and $\mathbf{x}^{j+}$ respectively in a $(k + 1)$-stage problem. It can easily be checked (using (2)) that $D_{ij}(\mathbf{x}, r) = 0$. Also, $D_{ij}(\mathbf{x}, h_k)$ and $D_{ij}(\mathbf{x}^{m-}, h_k)$ (for $m \in \{1, 2\}$) are non-positive by the inductive assumption. Hence, it suffices to show:

$$h_k(((\mathbf{x}^{i+})^{j+})^{a_1+}) - h_k((\mathbf{x}^{i+})^{a_0+}) - h_k((\mathbf{x}^{j+})^{b_1+}) + h_k(\mathbf{x}^{b_0+}) \ \le 0. \qquad (62)$$

One may proceed by considering all possible cases for the actions $a_1, b_0 \in \{0, 1, 2\}$. For each possible case, it suffices to find only one possible choice for a_0 and b_1 such that (62) holds. This is because a_0 and b_1 are defined as actions which *maximize* $h_k((\mathbf{x}^{i+})^{a_0+})$ and $h_k((\mathbf{x}^{j+})^{b_1+})$ respectively. Also, the number of cases to consider for (a_1, b_0) may be reduced by noting that the expression for $D_{ij}(\mathbf{x}, h_k)$ is symmetric in i and j (that is, $D_{ij}(\mathbf{x}, h_k)$=$D_{ji}(\mathbf{x}, h_k)$), so (for example) it is not necessary to treat $a_1 = b_0 = i$ and $a_1 = b_0 = j$ as separate cases.

After reductions, five possibilities remain for a_1 and b_0. These are considered below.

- If $a_1 = b_0 = 0$, one may consider $a_0 = b_1 = 0$, then (62) holds since the left-hand side is equal to $D_{ij}(\mathbf{x}, h_k) \le 0$.
- If $a_1 = b_0 = j$ for some $j \in \{1, 2\}$, then similarly we consider $a_0 = b_1 = j$, in which case (62) holds since the left-hand side is equal to $D_{ij}(\mathbf{x}^{j+}, h_k) \le 0$.
- If $a_1 = 0$ and $b_0 = j$ for some $j \in \{1, 2\}$, then we consider $a_0 = j$ and $b_1 = 0$, in which case (62) holds since the left-hand side is equal to zero.
- If $a_1 = j$ and $b_0 = i$ for some $i, j \in \{1, 2\}$ with $i \ne j$, we consider $a_0 = j$ and $b_1 = i$, in which case (62) holds since the left-hand side is equal to $D_{jj}(\mathbf{x}^{i+}, h_k) \le 0$.
- If $a_1 = j$ for some $j \in \{1, 2\}$ and $b_0 = 0$ then, since these actions attain the maximum in (58) at states $(\mathbf{x}^{i+})^{j+}$ and $\mathbf{x}$ respectively, it must be the case that $D_j((\mathbf{x}^{i+})^{j+}, h_k) \ge 0$ and $D_j(\mathbf{x}, h_k) \le 0$. However, the inductive assumptions of concavity and submodularity imply $D_j((\mathbf{x}^{i+})^{j+}, h_k) \le D_j(\mathbf{x}^{i+}, h_k) \le D_j(\mathbf{x}, h_k)$, so we conclude that $D_j((\mathbf{x}^{i+})^{j+}, h_k) = D_j(\mathbf{x}^{i+}, h_k) = D_j(\mathbf{x}, h_k) = 0$ in this case. By considering $a_0 = j$ and $b_1 = 0$, we find that the left-hand side of (62) simplifies to

$$D_j((\mathbf{x}^{i+})^{j+}, h_k) - D_j(\mathbf{x}, h_k),$$

which is equal to zero by our previous arguments.

The conclusion is that $D_{ij}(\mathbf{x}, h_{k+1})$ is non-positive in all possible cases, as required. Next, the diagonal submissiveness property will be considered. After some simplifications, one may derive the following expression for $D_{jj}(\mathbf{x}, h_{k+1})$:

$$\begin{aligned} &D_{jj}(\mathbf{x}, h_{k+1}) - D_{ij}(\mathbf{x}, h_{k+1}) = D_{jj}(\mathbf{x}, r) - D_{ij}(\mathbf{x}, r) \\ &\quad + \lambda\Delta\left(h_k(((\mathbf{x}^{j+})^{j+})^{a_1+}) - h_k(((\mathbf{x}^{i+})^{j+})^{a_0+}) - h_k((\mathbf{x}^{j+})^{b_1+}) + h_k((\mathbf{x}^{i+})^{b_0+})\right) \\ &\quad + \sum_{m=1}^{2} \min(x_m, 1)\mu_m\Delta\big(D_{jj}(\mathbf{x}^{m-}, h_k) - D_{ij}(\mathbf{x}^{m-}, h_k)\big) \\ &\quad + I(x_j = 0)\mu_j\Delta D_j(\mathbf{x}, h_k) + I(x_i = 0)\mu_i\Delta D_{jj}(\mathbf{x}, h_k) \\ &\quad + \left(1 - \lambda\Delta - \sum_{m=1}^{2} \min(x_m, 1)\mu_m\Delta - I(x_j = 0)\mu_j\Delta - I(x_i = 0)\mu_i\Delta\right)\big(D_{jj}(\mathbf{x}, h_k) - D_{ij}(\mathbf{x}, h_k)\big), \end{aligned} \tag{63}$$

where a_0, a_1, b_0 and b_1 are optimal actions at states $(\mathbf{x}^{i+})^{j+}$, $(\mathbf{x}^{j+})^{j+}$, $\mathbf{x}^{i+}$ and $\mathbf{x}^{j+}$ respectively in a $(k+1)$-stage problem. Using (2) one can verify that $D_{jj}(\mathbf{x}, r) - D_{ij}(\mathbf{x}, r) \leq 0$, where the inequality is strict only if $\mathbf{x}$ is a state with no customers present at facility j. In addition, $D_{jj}(\mathbf{x}, h_k) - D_{ij}(\mathbf{x}, h_k) \leq 0$ and $D_{jj}(\mathbf{x}^{m-}, h_k) - D_{ij}(\mathbf{x}^{m-}, h_k) \leq 0$ (for $m \in \{1, 2\}$) due to the inductive assumption. To continue the proof, it will be useful to show:

$$h_k(((\mathbf{x}^{j+})^{j+})^{a_1+}) - h_k(((\mathbf{x}^{i+})^{j+})^{a_0+}) - h_k((\mathbf{x}^{j+})^{b_1+}) + h_k((\mathbf{x}^{i+})^{b_0+}) \leq 0. \tag{64}$$

Again, this may be done by considering the possible cases for the actions $a_1, b_0 \in \{0, 1, 2\}$. This time, there are nine possibilities. These are considered below.

- If $a_1 = b_0 = 0$, then one may consider $a_0 = b_1 = 0$. Then (64) holds because the left-hand side is equal to $D_{jj}(\mathbf{x}, h_k) - D_{ij}(\mathbf{x}, h_k) \leq 0$.
- If $a_1 = b_0 = j$, then similarly we may consider $a_0 = b_1 = j$. Then (64) holds because the left-hand side is equal to $D_{jj}(\mathbf{x}^{j+}, h_k) - D_{ij}(\mathbf{x}^{j+}, h_k) \leq 0$.
- If $a_1 = b_0 = i$ then we consider $a_0 = b_1 = i$. Then (64) holds because the left-hand side is equal to $D_{jj}(\mathbf{x}^{i+}, h_k) - D_{ij}(\mathbf{x}^{i+}, h_k) \leq 0$.
- If $a_1 = i$ and $b_0 = j$ then we consider $a_0 = j$ and $b_1 = i$, in which case (64) holds trivially because the left-hand side is equal to zero.
- If $a_1 = 0$ and $b_0 = j$ then we consider $a_0 = 0$ and $b_1 = j$, in which case (64) again holds trivially since the left-hand side is equal to zero.
- If $a_1 = 0$ and $b_0 = i$ then we consider $a_0 = 0$ and $b_1 = i$ and aim to show:

$$h_k((\mathbf{x}^{j+})^{j+}) - h_k((\mathbf{x}^{i+})^{j+}) - h_k((\mathbf{x}^{i+})^{j+}) + h_k((\mathbf{x}^{i+})^{i+}) \leq 0. \tag{65}$$

Due to the inductive assumption, the left-hand side of (65) is bounded above by:

$$\begin{aligned} &h_k(\mathbf{x}^{j+}) - h_k(\mathbf{x}^{i+}) - h_k((\mathbf{x}^{i+})^{j+}) + h_k((\mathbf{x}^{i+})^{i+}) \\ &= h_k((\mathbf{x}^{i+})^{i+}) - h_k((\mathbf{x}^{i+})^{j+}) - h_k(\mathbf{x}^{i+}) + h_k(\mathbf{x}^{j+}), \end{aligned}$$

which is equal to $D_{ii}(\mathbf{x}, h_k) - D_{ij}(\mathbf{x}, h_k)$ and is therefore non-positive due to the same inductive assumption (with components i and j interchanged).

– If $a_1 = j$ and $b_0 = 0$, one may choose $a_0 = j$ and $b_1 = 0$. Then, due to the inductive assumption of diagonal submissiveness again, one may write:

$$\begin{aligned} &h_k(((\mathbf{x}^{j+})^{j+})^{j+}) - h_k(((\mathbf{x}^{i+})^{j+})^{j+}) - h_k(\mathbf{x}^{j+}) + h_k(\mathbf{x}^{i+}) \\ &\leq h_k((\mathbf{x}^{j+})^{j+}) - h_k((\mathbf{x}^{i+})^{j+}) - h_k(\mathbf{x}^{j+}) + h_k(\mathbf{x}^{i+}), \end{aligned}$$

which is equal to $D_{jj}(\mathbf{x}, h_k) - D_{ij}(\mathbf{x}, h_k) \leq 0$.

– If $a_1 = i$ and $b_0 = 0$ then we consider $a_0 = 0$ and $b_1 = i$. Then the left-hand side of (64) is equal to $D_{jj}(\mathbf{x}^{i+}, h_k)$, which is non-positive due to concavity of h_k.

– Finally, if $a_1 = j$ and $b_0 = i$, then the finite-stage optimality equations (58) imply $D_j((\mathbf{x}^{j+})^{j+}, h_k) \geq D_i((\mathbf{x}^{j+})^{j+}, h_k)$ and $D_j(\mathbf{x}^{i+}, h_k) \leq D_i(\mathbf{x}^{i+}, h_k)$. The former inequality implies $D_j(\mathbf{x}^{j+}, h_k) \geq D_i(\mathbf{x}^{j+}, h_k)$ due to the inductive assumption of diagonal submissiveness at stage k, which in turn implies $D_j(\mathbf{x}, h_k) \geq D_i(\mathbf{x}, h_k)$. By the same reasoning, the inequality $D_j(\mathbf{x}^{i+}, h_k) \leq D_i(\mathbf{x}^{i+}, h_k)$ implies $D_j(\mathbf{x}, h_k) \leq D_i(\mathbf{x}, h_k)$. Hence, $D_j(\mathbf{x}, h_k) \leq D_i(\mathbf{x}, h_k)$ and $D_j(\mathbf{x}, h_k) \geq D_i(\mathbf{x}, h_k)$ are both true, implying $D_j(\mathbf{x}, h_k) = D_i(\mathbf{x}, h_k) \geq 0$; that is, the actions i and j are both optimal at state $\mathbf{x}$. It has already been shown that (64) holds when $a_1 = b_0 = j$, so it must also hold when $a_1 = j$ and $b_0 = i$ given that $h_k(\mathbf{x}^{j+})$ and $h_k(\mathbf{x}^{i+})$ are equal.

These arguments confirm that (64) holds in all possible cases for the actions a_1 and b_0. It follows that the sum of the second, third and fifth lines on the right-hand side of (63) is non-positive. Hence, $D_{jj}(\mathbf{x}, h_{k+1}) - D_{ij}(\mathbf{x}, h_{k+1})$ may be bounded above as follows:

$$\begin{aligned} D_{jj}(\mathbf{x}, h_{k+1}) - D_{ij}(\mathbf{x}, h_{k+1}) &\leq D_{jj}(\mathbf{x}, r) - D_{ij}(\mathbf{x}, r) \\ &+ I(x_j = 0)\mu_j \Delta D_j(\mathbf{x}, h_k) + I(x_i = 0)\mu_i \Delta D_{jj}(\mathbf{x}, h_k). \end{aligned} \tag{66}$$

The term $I(x_i = 0)\mu_i \Delta D_{jj}(\mathbf{x}, h_k)$ is always non-positive due to the inductive assumption. If $x_j \geq 1$, then $I(x_j = 0) = 0$ and $D_{jj}(\mathbf{x}, r) - D_{ij}(\mathbf{x}, r) = 0$, so the right-hand side of (66) is non-positive as required. It remains only to consider the case $x_j = 0$. In this case, we will need to show that the following upper bound holds for all $n \in \mathbb{N}_0$:

$$D_j(\mathbf{x}, h_n) \leq \frac{\alpha_j \mu_j - \beta_j}{\mu_j \Delta}. \tag{67}$$

Indeed, property (67) may be established using a separate induction proof. It holds trivially when $n = 0$ since we define $v_0(\mathbf{x}) = 0$ for $\mathbf{x} \in S$ and $\alpha_j \mu_j - \beta_j$ is assumed positive in our model formulation. Proceeding by induction, we consider $n = k+1$ and write:

$$D_j(\mathbf{x}, h_{k+1}) = D_j(\mathbf{x}, r) + \lambda\Delta\left(h_k((\mathbf{x}^{j+})^{a_1+}) - h_k(\mathbf{x}^{a_0+})\right) + \sum_{i=1}^{2}\min(x_i, c_i)\mu_i\Delta D_j(\mathbf{x}^{i-}, h_k)$$
$$+\left(1 - \lambda\Delta - \sum_{i=1}^{2}\min(x_i, c_i)\mu_i\Delta - I(x_j = 0)\mu_j\Delta\right)D_j(\mathbf{x}, h_k), \tag{68}$$

where a_0 and a_1 are optimal actions at states $\mathbf{x}$ and $\mathbf{x}^{j+}$ respectively in a $(k+1)$-stage problem. Then, using the definition of a_0 as a *maximizing* action, we can write

$$h_k((\mathbf{x}^{j+})^{a_1+}) - h_k(\mathbf{x}^{a_0+}) \leq h_k((\mathbf{x}^{j+})^{a_1+}) - h_k(\mathbf{x}^{a_1+}) = D_j(\mathbf{x}^{a_1+}, h_k) \leq \frac{\alpha_j\mu_j - \beta_j}{\mu_j\Delta},$$

where the last inequality follows from the inductive assumption that (67) holds for $n = k$. Similarly, $D_j(\mathbf{x}^{i-}, h_k)$ is bounded above by $(\alpha_j\mu_j - \beta_j)/(\mu_j\Delta)$ for any $i \in \{1, 2, ..., N\}$, again due to the inductive assumption. To proceed, it will be convenient to consider the cases $x_j \geq 1$ and $x_j = 0$ separately. If $x_j \geq 1$, then $D_j(\mathbf{x}, r) = -\beta_j < 0$ and $I(x_j = 0) = 0$. Hence, due to the previous arguments, the right-hand side of (68) may be bounded above by a convex combination of terms bounded above by $(\alpha_j\mu_j - \beta_j)/(\mu_j\Delta)$, which establishes that $D_j(\mathbf{x}, h_{k+1}) \leq (\alpha_j\mu_j - \beta_j)/(\mu_j\Delta)$ as required. On the other hand, if $x_j = 0$, then $D_j(\mathbf{x}, r) = \alpha_j\mu_j - \beta_j > 0$. However, since $I(x_j = 0) = 1$ in this case, (68) implies (again making use of the inductive assumption):

$$D_j(\mathbf{x}, h_{k+1}) \leq \alpha_j\mu_j - \beta_j + (1 - \mu_j\Delta)\frac{\alpha_j\mu_j - \beta_j}{\mu_j\Delta} = \frac{\alpha_j\mu_j - \beta_j}{\mu_j\Delta}.$$

This completes the inductive proof that (67) holds for all $n \in \mathbb{N}_0$. Returning to the main part of our proof, we need to confirm that the left-hand side of (66) is non-positive in the case $x_j = 0$. In this case we have $I(x_j = 0) = 1$ and $D_{jj}(\mathbf{x}, r) - D_{ij}(\mathbf{x}, r) = -\alpha_j\mu_j$. Hence, using (67), we can write

$$D_{jj}(\mathbf{x}, h_{k+1}) - D_{ij}(\mathbf{x}, h_{k+1}) \leq -\alpha_j\mu_j + \mu_j\Delta\left(\frac{\alpha_j\mu_j - \beta_j}{\mu_j\Delta}\right) = -\beta_j < 0. \tag{69}$$

This confirms that $D_{jj}(\mathbf{x}, h_{k+1}) - D_{ij}(\mathbf{x}, h_{k+1}) \leq 0$ for all states $\mathbf{x} \in S$. Since it was shown earlier that $D_{ij}(\mathbf{x}, h_{k+1}) \leq 0$, it follows automatically that $D_{jj}(\mathbf{x}, h_{k+1}) \leq 0$, which completes the inductive proof that all three of the properties (59)-(61) hold for integers $n \in \mathbb{N}_0$. Since $h(\mathbf{x}) = \lim_{n\to\infty} h_n(\mathbf{x})$ for all $\mathbf{x} \in S$, the function h in (19) also shares these properties.

To complete the proof, we note that the existence of a monotone optimal policy can easily be proved if the function h satisfying the average reward optimality equations (19) is known to have the properties of concavity, submodularity and diagonal submissiveness. We simply need to define a policy

θ^* which always chooses an action attaining the maximum on the right-hand side of (19), and settles ties in a consistent way if two or more actions are tied. For example, suppose θ^* always prefers facility i to facility j if these two actions are tied, and never chooses to balk unless balking is the only action which attains the maximum. Then balking is chosen at state $\mathbf{x} \in S$ if and only if $D_i(\mathbf{x}, h)$ and $D_j(\mathbf{x}, h)$ are both negative. If this is the case, then we can use the concavity property to show $D_i(\mathbf{x}^{i+}, h)$ and $D_j(\mathbf{x}^{j+}, h)$ are also negative, while submodularity implies that $D_j(\mathbf{x}^{i+}, h)$ and $D_i(\mathbf{x}^{j+}, h)$ are also negative. Therefore θ^* must choose to balk at states $\mathbf{x}^{i+}$ and $\mathbf{x}^{j+}$, and property (a) in the statement of the theorem is satisfied. Similar arguments can be used to show that properties (b) and (c) are also satisfied and we omit the details here. □

F Proof of Theorem 4 (Statement 2).

In this proof we consider optimal policies in the single-facility case ($N = 1$). Let T_i denote the largest integer n such that an n-threshold policy is optimal at facility $i \in \{1, 2, ..., N\}$ given a demand rate $\lambda' > 0$. Our aim is to show that, given a demand rate $\lambda > \lambda'$ and an arbitrary integer $m > T_i$, it is impossible for an m-threshold policy to be optimal at facility i. We do this using the well-known technique of induction based on DP value iteration. For convenience, we will drop the facility index i from our notation since the facility is arbitrary. Given a demand rate $\lambda > 0$, let $v_k(x)$ denote the maximum achievable total reward in a finite-horizon problem consisting of k discrete time steps, given an initial state $x \in \mathbb{N}_0$, and let

$$h_k(x) := v_k(x) - v_k(0)$$

for all $k \in \mathbb{N}_0$ and $x \in \mathbb{N}_0$. In average reward problems, we must use *relative* value iteration in order to obtain finite values as $k \to \infty$ and this is accomplished by defining $h_k(x)$ as the relative benefit of starting in state x as opposed to starting at a fixed reference state (see Puterman (1994)); we use state 0 as the reference state here. The Bellman equation is

$$v_k(x) = r(x) + \max_{a \in \{0,1\}} \sum_{y \in \mathbb{N}_0} p(x, a, y) v_{k-1}(y) \quad (k \geq 1), \tag{70}$$

where the single-step rewards $r(x)$ and transition probabilities $p(x, a, y)$ are defined in (2) and (1) respectively (using scalars instead of vectors for the system state since we are considering a single facility). The corresponding finite-horizon values under demand rate $\lambda' < \lambda$ will be denoted by $v'_k(x)$ and $h'_k(x)$. Our aim is to show that, for all $k \in \mathbb{N}_0$ and $x \in \mathbb{N}_0$:

$$h'_k(x+1) - h'_k(x) \geq h_k(x+1) - h_k(x). \tag{71}$$

In order to complete the proof, we will also need to show the properties

$$h_k(x+1) - h_k(x) \leq \frac{\alpha\mu - \beta}{\mu \Delta} \tag{72}$$

and

$$h_k(x+2) - h_k(x+1) \le h_k(x+1) - h_k(x) \tag{73}$$

hold for $k \in \mathbb{N}_0$ and $x \in \mathbb{N}_0$. (Note that, since the demand rate λ is arbitrary, we can replace $h_k(\cdot)$ with $h'_k(\cdot)$ in (72) and (73) without loss of generality.) It will also simplify our notation to define

$$D(x, f) = f(x+1) - f(x)$$

for an arbitrary real-valued function f. We assume that $h_0(x) = h'_0(x) = 0$ for all $x \in \mathbb{N}_0$ and therefore (71), (72) and (73) hold trivially when $k = 0$. The principle of uniformization (see Serfozo (1979)) implies that, given a demand rate λ, we can choose the discrete-time step size to be any positive value less than or equal to $(\lambda + c_i\mu_i)^{-1}$. In particular, we can use the step size $(\lambda + c_i\mu_i)^{-1}$ under both of the demand rates λ and $\lambda' < \lambda$. We will therefore define the step size to be $\Delta = (\lambda + c_i\mu_i)^{-1}$ under both of these demand rates.

Our next task will be to show that (71), (72) and (73) hold with k replaced by $k+1$, under the assumption that they hold for an arbitrary $k \in \mathbb{N}_0$. We will begin with (72). By enumerating transition probabilities, we obtain

$$\begin{aligned} D(x, h_{k+1}) =& D(x, r) + \lambda\Delta(h_k(x+1+a_1) - h_k(x+a_0)) \\ &+ \min(x, c)\mu\Delta D(x-1, h_k) \\ &+ (1 - \lambda\Delta - \min(x, c)\mu\Delta - I(x < c)\mu\Delta)D(x, h_k), \end{aligned} \tag{74}$$

where I denotes the indicator function and $a_0, a_1 \in \{0, 1\}$ are actions which maximize $h_k(x+a_0)$ and $h_k(x+1+a_1)$ respectively. We note that, although the value function h_k only takes non-negative arguments, the term $\min(x, c)\mu\Delta D(x-1, h_k)$ can be ignored in the $x = 0$ case since $\min(0, c) = 0$.

We proceed by considering the possible cases for actions a_0 and a_1. If $a_0 = a_1 = 0$, then the expression

$$h_k(x+1+a_1) - h_k(x+a_0) \tag{75}$$

is equal to $D(x, h_k)$, which is bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$ by the inductive assumption. Similarly, if $a_0 = a_1 = 1$ then (75) is equal to $D(x+1, h_k)$, which is also bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$. If $a_0 = 1$ and $a_1 = 0$ then (75) is trivially equal to zero. Finally, if $a_0 = 0$ and $a_1 = 1$ then this implies that the inequalities

$$D(x, h_k) \le 0, \quad D(x+1, h_k) \ge 0$$

both hold. However, property (73) tells us that $D(x+1, h_k) \le D(x, h_k)$ and therefore the above inequalities imply $D(x+1, h_k) = D(x, h_k) = 0$. With $a_0 = 0$ and $a_1 = 1$, (75) is equal to $D(x+1, h_k) + D(x, h_k) = 0$. Thus, we have shown that in all possible cases for a_0 and a_1, the difference (75) is bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$.

We still need to consider the other terms on the right-hand side of (74). We will consider the cases $x < c$ and $x \ge c$ separately. If $x < c$, then

$$\begin{aligned} D(x, h_{k+1}) =& D(x, r) + \lambda\Delta(h_k(x+1+a_1) - h_k(x+a_0)) \\ &+ x\mu\Delta D(x-1, h_k) + (1 - \lambda\Delta - (x+1)\mu\Delta)D(x, h_k). \end{aligned}$$

By the inductive assumption $D(x-1, h_k)$ and $D(x, h_k)$ are bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$, and we have shown that the same is true of $h_k(x+1+a_1) - h_k(x+a_0)$. Therefore, noting that $D(x,r) = \alpha\mu - \beta$ for $x < c$, we have

$$D(x, h_{k+1}) \leq \alpha\mu - \beta + (1-\mu\Delta)(\alpha\mu - \beta)/(\mu\Delta) = (\alpha\mu - \beta)/(\mu\Delta)$$

as required. In the $x \geq c$ case, we have

$$\begin{aligned} D(x, h_{k+1}) =& D(x, r) + \lambda\Delta(h_k(x+1+a_1) - h_k(x+a_0)) \\ &+ c\mu\Delta D(x-1, h_k) + (1 - \lambda\Delta - c\mu\Delta)D(x, h_k). \end{aligned} \tag{76}$$

The last three terms on the right-hand side of (76) represent a convex combination of terms bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$. Also, given that $x \geq c$, we have $D(x,r) = -\beta < 0$ and therefore $D(x, h_{k+1})$ is bounded above by $(\alpha\mu - \beta)/(\mu\Delta)$ as required. This completes the proof that (72) holds with k replaced by $k+1$.

Next, we consider (73). Enumerating the transition probabilities again, we obtain (after some simplifications):

$$\begin{aligned} &D(x+1, h_{k+1}) - D(x, h_{k+1}) = D(x+1, r) - D(x, r) \\ &+ \lambda\Delta([h_k(x+2+a_2) - h_k(x+1+a_1)] - [h_k(x+1+a_1) - h_k(x+a_0)]) \\ &+ \min(x,c)\mu\Delta(D(x, h_k) - D(x-1, h_k)) \\ &+ I(x = c-1)\mu\Delta D(x, h_k) \\ &+ (1 - \lambda\Delta - \min(x,c)\mu\Delta - I(x < c-1)\mu\Delta - I(x < c)\mu\Delta)(D(x+1, h_k) - D(x, h_k)), \end{aligned} \tag{77}$$

where $a_0, a_1, a_2 \in \{0,1\}$ are actions which maximize $h_k(x+a_0)$, $h_k(x+1+a_1)$ and $h_k(x+2+a_2)$ respectively. The reward function $r(x)$ increases linearly (with gradient $\alpha\mu - \beta > 0$) when $x < c$, and decreases linearly (with gradient $-\beta < 0$) when $x \geq c$. Hence:

$$D(x+1, r) - D(x, r) = \begin{cases} -\alpha\mu, & \text{if } x = c-1, \\ 0, & \text{otherwise,} \end{cases}$$

so $D(x+1, r) - D(x, r)$ is non-positive. Next, we will show that the second line on the right-hand side of (77) is non-positive by considering the possible combinations for a_0, a_1 and a_2. These are considered below.

- If $a_0 = a_2 = 0$, then

  $$[h_k(x+2+a_2) - h_k(x+1+a_1)] - [h_k(x+1+a_1) - h_k(x+a_0)] \tag{78}$$

  is bounded above by $D(x+1, h_k) - D(x, h_k)$. This is due to the fact that $a_1 \in \{0,1\}$ is a *maximizing* action, so the value of (78) must be no greater than the value we would obtain by putting $a_1 = 0$. Since $D(x+1, h_k) - D(x, h_k) \leq 0$ by the inductive assumption, we can say that (78) is non-positive.

- If $a_0 = a_2 = 1$, then we can use similar reasoning to the case above and say that (78) is bounded above by $D(x+2, h_k) - D(x+1, h_k)$ (by considering $a_1 = 1$), and this is also non-positive by the inductive assumption.
- If $a_0 = 1$ and $a_2 = 0$, then (78) is bounded above by

$$h_k(x+2) - h_k(x+2) - h_k(x+1) + h_k(x+1) = 0.$$

- Finally, if $a_0 = 0$ and $a_2 = 1$, then (78) is bounded above by

$$h_k(x+3) - h_k(x+2) - h_k(x+1) + h_k(x),$$

  which is equivalent to $D(x+2, h_k) - D(x, h_k)$. By the inductive assumption we can say that $D(x+2, h_k) \leq D(x+1, h_k) \leq D(x, h_k)$, so (78) must again be non-positive.

By the inductive assumption, the differences $D(x, h_k) - D(x-1, h_k)$ and $D(x+1, h_k) - D(x, h_k)$ appearing on the third and fifth lines of (77) respectively are non-positive. On the fourth line we have an expression which is non-zero only if $x = c-1$, in which case it is bounded above by $\alpha\mu - \beta$ due to property (72). However, in the $x = c-1$ case we also have $D(x+1, r) - D(x, r) = -\alpha\mu$, so we can conclude that for any $x \in \mathbb{N}_0$:

$$D(x+1, r) - D(x, r) + I(x = c-1)\mu\Delta D(x, h_k) \leq 0.$$

This completes the proof that the sum of terms on the right-hand side of (77) is non-positive, as required. At this point we have proved that (72) and (73) hold with k replaced by $k+1$, and next we turn our attention to (71). By again enumerating transition probabilities, we have

$$\begin{aligned} D(x, h'_k) - D(x, h_k) =& \lambda'\Delta([h'_k(x+1+a'_1) - h'_k(x+a'_0)] - [h_k(x+1+a_1) - h_k(x+a_0)]) \\ &+ (\lambda - \lambda')\Delta(D(x, h'_k) - [h_k(x+1+a_1) - h_k(x+a_0)]) \\ &+ \min(x, c)\mu\Delta(D(x-1, h'_k) - D(x-1, h_k)) \\ &+ (1 - \lambda\Delta - \min(x, c)\mu\Delta - I(x < c)\mu\Delta)(D(x, h'_k) - D(x, h_k)), \end{aligned} \tag{79}$$

where a_0, a_1, a'_0 and a'_1 are all actions (belonging to $\{0, 1\}$) which maximize the relevant quantities $h_k(x+a_0)$, $h_k(x+1+a_1)$, etc. By induction we assume that the differences $D(x-1, h'_k) - D(x-1, h_k)$ and $D(x, h'_k) - D(x, h_k)$ are both non-negative and therefore only the first and second lines on the right-hand side of (79) remain to be checked for non-negativity. We need to show that, for each possible combination of actions $a_0, a_1, a'_0, a'_1 \in \{0, 1\}$, we have

$$\begin{aligned} &\lambda'\Delta([h'_k(x+1+a'_1) - h'_k(x+a'_0)] - [h_k(x+1+a_1) - h_k(x+a_0)]) \\ &\quad + (\lambda - \lambda')\Delta(D(x, h'_k) - [h_k(x+1+a_1) - h_k(x+a_0)]) \\ &\geq 0. \end{aligned} \tag{80}$$

In practice, since a_0, a_1, a'_0, a'_1 are all *maximizing* actions, we only need to consider the possible cases for a'_0 and a_1 and show (in each case) that there exists some choice of $(a'_1, a_0) \in \{0, 1\}^2$ such that (80) holds. The possible cases are considered below.

- If $a'_0 = a_1 = 0$, then we consider $a'_1 = a_0 = 0$ also, in which case (80) holds by the inductive assumption (71).
- If $a'_0 = a_1 = 1$, then we consider $a'_1 = a_0 = 1$, in which case the first term on the left-hand side of (80) is non-negative by the inductive assumption, and the second term simplifies to

$$(\lambda - \lambda')\Delta(D(x, h'_k) - D(x+1, h_k)).$$

Due to property (73), the above expression is bounded below by

$$(\lambda - \lambda')\Delta(D(x, h'_k) - D(x, h_k)),$$

which is non-negative due to (72).
- If $a'_0 = 1$ and $a_1 = 0$, then we consider $a'_1 = 0$ and $a_0 = 1$. Then the first term on the left-hand side of (80) is equal to zero, and the second term simplifies to $(\lambda - \lambda')\Delta D(x, h'_k)$. Given that $a'_0 = 1$, it must be the case that $D(x, h'_k) \geq 0$, so (80) holds as required.
- Finally, if $a'_0 = 0$ and $a_1 = 1$, this implies $D(x, h'_k) \leq 0$ and $D(x + 1, h_k) \geq 0$. The latter inequality implies $D(x, h_k) \geq 0$ due to (73), but given that $D(x, h'_k) \geq D(x, h_k)$ by the inductive assumption we then obtain $D(x, h_k) = D(x, h'_k) = D(x+1, h_k) = 0$. By considering $a'_1 = 0$ and $a_0 = 1$ in (80) the left-hand side simplifies to

$$\lambda'\Delta(D(x, h'_k) - D(x+1, h_k)) + (\lambda - \lambda')\Delta(D(x, h'_k) - D(x+1, h_k)),$$

which is equal to zero by our previous arguments.

This completes the proof that (71) holds with k replaced by $k + 1$. At this point we have shown that (71), (72) and (73) hold for all $k \in \mathbb{N}_0$ and $x \in \mathbb{N}_0$. As $k \to \infty$, h_k tends to a function h which satisfies the average reward optimality equations (see Puterman (1994)). As such, let h and h' denote the functions satisfying the optimality equations under demand rates λ and $\lambda' < \lambda$ respectively; that is, $h = \lim_{k\to\infty} h_k$ and $h' = \lim_{k\to\infty} h'_k$. Given that T is the largest integer n such that an n-threshold policy is optimal under a demand rate $\lambda' > 0$, the optimality equations imply that

$$h'(m + 1) - h'(m) < 0 \quad \forall m \geq T.$$

Our inductive proof implies that $h(x + 1) - h(x) \leq h'(x + 1) - h'(x)$ for any $x \in \mathbb{N}_0$, so from the above we obtain $h(m + 1) - h(m) < 0$ for $m \geq T$, which proves that T is the greatest possible optimal threshold under the larger demand rate λ. Recalling that an arbitrary facility $i \in \{1, 2, ..., N\}$ was considered throughout this proof, it follows that if balking is chosen by the Whittle policy at state $\mathbf{x} \in S$ under demand rate λ' then it must also be chosen at $\mathbf{x}$ under demand rate λ, and so we obtain $S_W(\lambda) \subseteq S_W(\lambda')$ as stated by the theorem. □

G Proof of Theorem 7 and a counter-example to show lack of heavy-traffic optimality for $\theta^{[B]}$.

First, we will prove Theorem 7. Let $\Lambda = (\lambda_1, ..., \lambda_N)$ be any static routing policy under which the system is stable; that is, $\lambda_i < c_i\mu_i$ for each $i \in \{1, ..., N\}$. The long-run average reward $g^{\Lambda}(\lambda)$ under this policy is given by (20). By Little's formula (see Gross and Harris (1998), p.10) we have $L_i(\lambda_i) = \lambda_i M_i(\lambda_i)$ for each facility i, where $M_i(\lambda_i)$ is an individual customer's average waiting time in the system. Hence, we can also write

$$g^{\Lambda}(\lambda) = \sum_{i=1}^{N} \lambda_i(\alpha_i - \beta_i M_i(\lambda_i)). \tag{81}$$

As the system demand rate λ tends to zero, the constraint $\sum_{i=1}^{N} \lambda_i \leq \lambda$ implies that $\lambda_i \to 0$ for each $i \in \{1, ..., N\}$, which in turn implies that $M_i(\lambda_i) \to 1/\mu_i$ (equivalently, a customer's expected waiting time at facility i tends to their expected service time). Let J denote the set of facilities which maximize $\alpha_i - \beta_i/\mu_i$. That is:

$$J = \arg\max_{i \in \{1,...,N\}} \left\{ \alpha_i - \frac{\beta_i}{\mu_i} \right\}.$$

Also, let $\bar{\Lambda} = (\bar{\lambda}_1, ..., \bar{\lambda}_N)$ be a static policy which, given any demand rate $\lambda > 0$, always chooses to join a particular facility $j \in J$ with probability one; that is, $\bar{\lambda}_j = \lambda$ for some $j \in J$ and $\bar{\lambda}_i = 0$ for all $i \neq j$. Since we are considering a light-traffic limit, we can assume that $\lambda < c_j\mu_j$ and hence the system is stable under $\bar{\Lambda}$. Then, by our previous arguments:

$$\lim_{\lambda \to 0} \frac{g^{\bar{\Lambda}}(\lambda)}{\lambda} = \lim_{\lambda \to 0} \sum_{i=1}^{N} \frac{\bar{\lambda}_i}{\lambda}(\alpha_i - \beta_i M_i(\lambda_i)) = \max_{i \in \{1,...,N\}} \left\{ \alpha_i - \frac{\beta_i}{\mu_i} \right\}. \tag{82}$$

Meanwhile, it is clear that an upper bound for the expected long-run average reward under *any* admissible policy (static or otherwise) is $\lambda \max_i\{\alpha_i - \beta_i/\mu_i\}$, since an individual customer's expected net reward can never be greater than $\max_i\{\alpha_i - \beta_i/\mu_i\}$ (see 4). Therefore the equation (82) implies that the static policy $\bar{\Lambda}$ achieves asymptotic light-traffic optimality, and so this must also be the case for the optimal static policy Λ^*.

Next, consider the Bernoulli improvement policy $\theta^{[B]}$. The method of construction for this policy implies that, given any demand rate $\lambda > 0$, it will improve upon the expected long-run average reward earned by the optimal static policy Λ^*. Indeed, the action chosen by $\theta^{[B]}$ at state $\mathbf{x}$ is that which maximizes the expression $\delta(\mathbf{x}, a)$ defined in (21). Noting that

$$\sum_{i=1}^{N} \frac{\lambda_i^*}{\lambda}\delta(\mathbf{x}, i) + \frac{\lambda_0^*}{\lambda}\delta(\mathbf{x}, 0) = 0,$$

it must be the case that $\delta(\mathbf{x}, a) \geq 0$ for at least one action $a \in \{0, 1, ..., N\}$. It follows from the definition of $\delta(\mathbf{x}, a)$ that $g^{[B]}(\lambda) \geq g^{\Lambda^*}(\lambda)$ for all $\lambda > 0$ and therefore $\theta^{[B]}$ inherits the light-traffic optimality of Λ^*. □

Next, we provide a counter-example to show that $\theta^{[B]}$ is not necessarily optimal in a heavy-traffic limit. We consider a single facility ($N = 1$) and a demand rate $\lambda = 5$. The parameters for the single facility are given as

$$c = 5, \quad \mu = 2, \quad \beta = 15, \quad \alpha = 8.$$

Let λ_1^* be the Poisson demand rate for the facility under the optimal static policy. Solving the relevant convex optimization problem yields $\lambda_1^* \approx 3.595$, which gives an average reward of approximately 1.458. We also note that, due to the concavity of the objective function, the same optimal value λ_1^* will be obtained for any value of λ greater than 5.

It can be seen from (28) and (30) that the index $D_i(x, \lambda_1^*)$ is independent of the system demand rate λ. In this example, $D_i(x, \lambda_1^*)$ is negative if and only if $x \geq 3$. Hence, for all demand rates $\lambda \geq 5$, the Bernoulli improvement policy $\theta^{[B]}(\lambda)$ is a threshold policy with a threshold of 3. Note that if a single facility is operating under a threshold policy and $T \in \mathbb{N}$ denotes the threshold, then the long-run average reward under the threshold policy tends to

$$\min(c, T)\alpha\mu - \beta T$$

as the demand rate tends to infinity. Hence, in this example, we have

$$\lim_{\lambda \to \infty} g^{[B]}(\lambda) = 3(\alpha\mu - \beta) = 3.$$

However, by increasing the threshold to 5, we would obtain a policy under which the average reward tends to $5(\alpha\mu - \beta) = 5$ as $\lambda \to \infty$. Hence, for sufficiently large λ, we find that $\theta^{[B]}(\lambda)$ is sub-optimal.

H Methods for generating the parameters for the numerical experiments in Section 6

For the 32,934 systems considered in Section 6.1, the parameters were randomly generated as follows:

- The number of facilities, N, was sampled unbiasedly from the set $\{2, 3, 4\}$.
- Each service rate μ_i was sampled from a uniform distribution between 5 and 25.
- Each service capacity c_i was sampled unbiasedly from the set $\{1, 2, 3, 4, 5\}$.
- Each holding cost β_i was sampled from a uniform distribution between 5 and 25.
- Each fixed reward α_i was sampled from a uniform distribution which was dependent upon the number of facilities N. This uniform distribution was between 2 and 28 in the case $N = 2$, between 2 and 18 in the case $N = 3$, and between 2 and 8 in the case $N = 4$.

– The demand rate λ was sampled from a uniform distribution between 0 and $1.5 \times \sum_{i=1}^{N} c_i\mu_i$.

In addition, all facilities i were required to satisfy the condition $\alpha_i > \beta_i/\mu_i$ in order to avoid degeneracy, and the size of $\tilde{S}$ (defined in (5)) was required to be between 100 and 100,000 (inclusive).

Parameter sets which did not satisfy these criteria were rejected and, in these cases, all parameters were re-sampled. (Note: as a result of the requirement for $\tilde{S}$ to be between 100 and 100,000, parameter sets with $N = 2$ were somewhat more likely to be accepted than those with $N = 3$ or $N = 4$. This can be seen in Section 6 (see Table 2).)

For the 4660 systems considered in Section 6.2, parameters were randomly generated as follows:

– The number of facilities, N, was sampled unbiasedly from the set $\{4, 5, 6, 7, 8, 9, 10, 11, 12\}$.
– Each service capacity c_i was sampled unbiasedly from the set $\{2, 3, 4, 5, 6\}$.
– Each fixed reward α_i was sampled from a uniform distribution between 2 and 20.
– The parameters μ_i and β_i and the demand rate λ were generated in the same way as in Section 6.1.

As in Section 6.1, all facilities i were required to satisfy the condition $\alpha_i > \beta_i/\mu_i$. For these 'large system' experiments, the size of $\tilde{S}$ was required to be greater than 100,000. However, in order to place a restriction on the number of indices to be computed, we also enforced the requirement that the selfish 'caps' for the individual facilities, given by $\lfloor \alpha_i c_i \mu_i / \beta_i \rfloor$ (see (5)) should not exceed 100. Parameter sets which did not satisfy the aforementioned criteria were rejected and, in these cases, all parameters were re-sampled.