MANAGEMENT SCIENCE
Vol. 60, No. 1, January 2014, pp. 110–129
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.2013.1754

Optimal Hiring and Retention Policies for Heterogeneous Workers Who Learn

Alessandro Arlotto
The Fuqua School of Business, Duke University, Durham, North Carolina 27708,
aa249@duke.edu

Stephen E. Chick
Technology and Operations Management Area, INSEAD, 77305 Fontainebleau France,
stephen.chick@insead.edu

Noah Gans
Operations and Information Management Department, The Wharton School, University of Pennsylvania,
Philadelphia, Pennsylvania 19104, gans@wharton.upenn.edu

We study the hiring and retention of heterogeneous workers who learn over time. We show that the problem can be analyzed as an infinite-armed bandit with switching costs, and we apply results from Bergemann and Välimäki [Bergemann D, Välimäki J (2001) Stationary multi-choice bandit problems. *J. Econom. Dynam. Control* 25(10):1585–1594] to characterize the optimal hiring and retention policy. For problems with Gaussian data, we develop approximations that allow the efficient implementation of the optimal policy and the evaluation of its performance. Our numerical examples demonstrate that the value of active monitoring and screening of employees can be substantial.

Key words: learning curves; heterogeneous workers; Bayesian learning; call center; hiring and retention; operations management; Gittins index; Bandit problem
History: Received August 23, 2010; accepted March 14, 2013, by Yossi Aviv, operations management. Published online in *Articles in Advance* September 27, 2013.

Workers can have diverse capabilities that change through time. In many settings, such as call centers and manufacturing, on-the-job learning is an important element of operational performance. Learning can take a number of forms, including decreases in the time required to complete tasks and improvements in quality. Employee turnover can similarly affect organizational performance. Workers who turn over (quit) or are terminated may be replaced by new hires who differ in both ability and experience. Different policies for hiring, monitoring, and retaining employees will influence the long-run performance of a firm.

Often there can be uncertainty regarding employee capabilities. Significant random variations in task times or quality—driven by task-by-task variability—can make it difficult for an employer to infer a given employee's efficiency or quality, particularly for new employees who have little or no previous track record.

Uncertainty, together with these many sources of variation—across employee capabilities, across tasks, and over time—makes decisions regarding the retention of workers complex. The longer a worker is retained, the better the inference an employer can make regarding his or her capabilities. On-the-job learning, which can lead to quality improvements in incumbent employees, also favors employee retention. However, the opportunity cost of retaining a poor performer can be great, particularly if there is wide variation in quality across the population of potential hires.

In this paper, we develop and analyze a model that integrates all of these factors. In our model, an employer (referred to as "she") seeks to hire and retain a fixed number of employees from an infinite, heterogeneous population of potential hires. Each employee (referred to as "he") repeatedly performs the same task, whose cost the employer wishes to minimize or, equivalently, whose quality is to be maximized. Each hire moves down a learning curve, but elements of the curve's parameters are unknown to the employer. The employer takes a Bayesian view of employees' types. By repeatedly observing the task performance of a given worker, she can make increasingly better judgments concerning his quality. After each such task, the employee decides whether or not he wants to continue working. Given that the worker decides to stay, the employer can decide whether to retain him or to replace him with a new hire. Each of these decisions has a cost for the employer. A quitting cost is incurred when a worker quits, a switching cost is incurred when a worker is terminated, and a training cost is incurred for each newly hired employee.

We formulate this problem as an infinite-horizon, discounted problem in which, at any time, the employer uses a single worker, and we show that the problem can be modeled as a multiarmed bandit problem with switching costs and an infinite number of arms. We then apply well-known results, developed by Bergemann and Välimäki (2001), to characterize the optimal hiring and retention policy and find that a Gittens-index policy is optimal. Furthermore, the optimal policy exhibits a "no-recall" property that is useful from an application perspective. (Farias and Madan 2011 analyze no-recall policies for finite-armed bandits.) These Gittins-index results extend to more complex settings, including contexts with multiple employees and environments with multiple, heterogeneous pools of potential employees.

For specific common forms of the learning-curve function, we delineate a simple stopping boundary and then use the boundary to develop approximations to the Gittins index that are straightforward to calculate and implement. These approximations are then the basis of numerical examples.

Our numerical results provide insights into the nature and performance of the optimal policy. They show how the stopping boundary reflects a tradeoff between two types of learning: the performance improvement that is linked to an employee's on-the-job experience, and the statistical learning that allows the employer to make better judgments concerning a worker's ability. They demonstrate that the value of active monitoring and screening of employees can be substantial. They reveal that the early stages of workers' tenures are the most important for the effectiveness of the optimal policy and, in turn, suggest simpler hiring policies that have the potential to perform well, within a few percent of optimality.

Sensitivity analysis with respect to model parameters provides further insights. In addition to direct gains that accrue from steeper learning curves, investments in employee learning can provide an important secondary benefit: the optimality of lower termination rates. Reductions in the variability of task performance can improve the sensitivity of screening procedures and similarly reduce optimal termination rates. The ability to terminate employees should motivate managers to consider a broader spectrum of potential hires.

1. Literature Review

There is a vast empirical literature on learning-curve phenomena (Yelle 1979), as well as papers devoted to effective managerial control of factors that affect or depend on learning (Dada and Srikanth 1990, Wiersma 2007). Much of it is segmented into the individual (e.g., Nembhard and Uzumeri 2000a, Nembhard 2001) and organizational levels (e.g., Bailey 1989, Lapré et al. 2000, Pisano et al. 2001). Nembhard and Uzumeri (2000b) provide a unified study that considers both. Our analysis focuses on the individual level.

The literature that explicitly addresses both worker heterogeneity and learning is much smaller. Most closely related to our work is Nagypál (2007), which models both learning about match quality (between workers and a firm) and learning by doing. That paper's aims and results differ significantly from ours. Its model and analysis enable the use of statistical methods to discriminate between the two forms of learning in empirical employment records. We focus on model-based, and normative insights into the nature of effective retention/termination decisions.

A few recent papers in operations-related fields also address dimensions of heterogeneity in learning and employee retention. Shafer et al. (2001) provide empirical evidence of the heterogeneity of learning curves across individuals who assemble car radios. Pisano et al. (2001) document heterogeneity across hospital units that perform cardiac surgery. Mazzola and McCardle (1996, 1997) develop models to estimate uncertain learning curves and to control production run lengths, given that a firm faces this uncertainty. None of these papers consider uncertainty regarding learning curves across individuals or groups, however. Neither do they address employee turnover or employee retention decisions.

Shafer et al. (2001) consider individual learning curves and show that, by not considering learning-parameter variations across workers, one may significantly underestimate overall productivity, given workers who operate independently. Nembhard and Osothsilp (2002) show how task complexity affects the distribution of individual learning and forgetting parameters. Gans et al. (2010) show that the service times of call center agents reflect on-the-job learning, as well as agent heterogeneity.

There also exists a rich literature that addresses labor quality and selection. The literature on secretary problems develops a normative approach to the initial screening and hiring of employees who come from a heterogeneous pool (Freeman 1983). Similarly, there is work on multiarmed bandit problems that addresses matching problems in labor markets: typically, problems in which employees choose firms (Jovanovic 1979, Sundaram 2005).

In our context, this work can be reinterpreted as addressing firms choosing employees, and we use results concerning infinite-armed bandits with switching costs (but no learning) to characterize optimal hiring and retention policies (Banks and Sundaram 1992, Bergemann and Välimäki 2001, Sundaram 2005). The most closely related work on (finite-armed) bandit models with switching costs can be found in Weitzman (1979), Banks and Sundaram (1994), Asawa and Teneketzis (1996), Jun (2004), and Niño-Mora (2008).

The managerial implications of learning have received less attention. Nembhard (2001) is the first to propose a method that assigns workers to tasks based on learning rates of individuals, considers forgetting as well as learning, and offers heuristics for managers. Our work differs in that we derive optimal policies and our numerical experiments use somewhat different learning curves.

Pinker and Shumsky (2000), Gans and Zhou (2002), and Whitt (2006) study learning with respect to the operations management/human resource management (OM/HRM) interface. Their work does not take into account worker heterogeneity. Gans et al. (2003) and Akşin et al. (2007) are recent surveys that include discussion of learning and HRM in the call center industry. Gaimon (1997) and Carillo and Gaimon (2000) study the importance of learning when new technologies are introduced. Gaimon et al. (2011) use mathematical models and empirical data to assess learning before doing, which can be modeled as training costs in our analysis, and learning by doing, which is modeled by learning curves. Goldberg and Touw (2003) consider statistical inference of learning curve parameters in a managerial context.

2. The Hiring and Retention Problem with One Employee

In this section, we define the problem of an employer who requires the services of a single worker and who, at each discrete period of time, decides whether to retain the current employee or to terminate him and hire someone else from an infinite pool of workers. The assumption that there exists an infinite pool of potential hires is appropriate in so-called employers' markets, in which the potential workforce is sufficiently large that workers who quit need not be considered again. Section 4 explores the employment of multiple hires, as well as the presence of several heterogeneous pools of workers.

At each time $t = 0, 1, 2, \ldots$, the employer requires the service of a single employee, i, drawn from an infinite pool of potential workers, $\mathcal{S}_t$; $\mathcal{S}_0$ represents the initial pool from which the employer can draw. If employee i quits at time t, then he is removed from the pool of potential hires and $\mathcal{S}_{t+1} = \mathcal{S}_t \backslash \{i\}$. We let $\pi(t) = i \in \mathcal{S}_t$ denote the employer's choice of employee i at time t and define $\pi = \{\pi(0), \pi(1), \ldots\}$ to be a *hiring and retention policy* that specifies which workers the employer engages over time.

The performance of potential workers is uncertain and evolving over time. If worker $i \in \mathcal{S}_t$ is employed at time t, then his performance is defined by the following relation:

$$Z_{i,t} = g(\boldsymbol{\theta}_i, n_{i,t}, \epsilon_{i,t}), \tag{1}$$

where $\boldsymbol{\theta}_i \in \Omega$ is a vector of parameters that reflects worker i's ability; $n_{i,t} = 0, 1, 2, \ldots$ reflects his experience to date; $\epsilon_{i,t}$ is a noise term with support $\mathcal{E}$; and $g(\cdot)$ is a deterministic function of its arguments. We denote the realization of $Z_{i,t}$ by $z_{i,t}$. For $\boldsymbol{\theta}_i = (a_i, b_i)$, Yelle (1979) describes the following commonly used form:

$$Z_{i,t} = \exp(a_i + b_i \ln(n_{i,t} + 1) + \epsilon_{i,t}), \qquad n_{i,t} = 0, 1, 2, \ldots. \tag{2}$$

Here, a_i is a parameter that determines a base level of performance and $b_i < 0$ describes the rate of learning. If $Z_{i,t}$ were task time, then a_i and b_i would be scaled in the logarithm of the time unit.

The structural results concerning optimal policies, in §3, require only the general functional form (1), together with some technical assumptions. Furthermore, the function $g(\cdot)$ is quite general and, in addition to learning, might reflect the effect of other factors such as fatigue. Although our analysis does hinge on a single measure of performance, the representation of an outcome, $Z_{i,t}$, can be generalized to explicitly represent multiple dimensions (such as revenue, cost, quality) that are aggregated into a single score by using a functional. Section 5, in which we develop methods for explicitly calculating the stopping boundaries necessary to implement optimal policies, assumes a more specific form of $Z_{i,t}$, such as that given by (2).

At the end of a given period, after his performance, the current employee notifies the employer of his intention to continue working or to leave. So, we associate with each worker a sequence of Bernoulli leaving decisions, $\mathbf{L}_i = (L_{i,0}, L_{i,1}, L_{i,2}, \ldots)$, indexed only by experience, such that worker i leaves or quits at the end of period t, after his $(n_{i,t}+1)$st performance, if and only if $L_{i,0} = L_{i,1} = \cdots = L_{i,n_{i,t}-1} = 0$ and $L_{i,n_{i,t}} = 1$. We denote the realization of $\mathbf{L}_i$ and $L_{i,n_{i,t}}$ by $\mathbf{l}_i$ and $l_{i,n_{i,t}}$, respectively. In this paper, we alternatively use $\mathbb{1}(E)$ or $\mathbb{1}_E$ to denote the indicator function of the event E and, for any hiring policy π and for each worker $i \in \mathcal{S}_0$, we let

$$\Lambda_i(\pi) = \sum_{t=0}^{\infty} \mathbb{1}(\pi(t) = i) \tag{3}$$

be i's working lifetime: the number of periods he is employed. In turn, we define worker i's *quitting* probability, $q_{i,n}$, to be

$$q_{i,n} = \mathbb{P}(L_{i,n} = 1 \mid \Lambda_i(\pi) \geq n+1), \tag{4}$$

and call $1 - q_{i,n}$ worker i's *continuation* probability.

For $t \geq 0$, let $\mathscr{H}_{i,t} = \{(z_{\pi(s),s}, l_{\pi(s),n_{\pi(s),s}})\colon \pi(s) = i, s \leq t\}$ ($\mathscr{H}_{i,0} = \varnothing$) denote worker i's *employment history* up to time t. The quitting probability of an employee with experience $n_{i,t}$, $q_{i,n_{i,t}}$, may depend on $\mathscr{H}_{i,t}$ and on his ability $\boldsymbol{\theta}_i$, but it is assumed to be independent of the employer's hiring policy, π:

$$\mathbb{P}(L_{i,n}=1 \mid \Lambda_i(\pi) \geq n+1) = \mathbb{P}(L_{i,n}=1 \mid \Lambda_i(\pi') \geq n+1) \quad \text{for all } \pi \neq \pi' \text{ and all } i, n.$$

This independence assumption is restrictive, and it is not difficult to imagine how employee turnover decisions may be influenced by the employer's retention (and compensation) policies. For example, by paying better performers more, the employer could provide an incentive for employee turnover patterns to change in a manner that is favorable to her. The inclusion of these types of incentives and responses extends the analysis of the employer's hiring and retention problem from the realm of single-decision-maker optimization problems to that of stochastic games and is beyond the focus of our current work. Nevertheless, the strategic interaction of employer and employees is both interesting and important, and we will briefly return to this issue in the numerical results of §6.

The employer does not know each employee's $\boldsymbol{\theta}_i$ or $\mathbf{l}_i$ in advance. Rather, she believes that there exists a random vector, $\boldsymbol{\Theta}$, that reflects the distribution of abilities in the population of potential workers, and a random set of leaving decisions, $\mathbf{L}$. The distributions for $\boldsymbol{\Theta}$ and $\mathbf{L}$ can be estimated using historical data and statistical techniques.

Each time the employer hires a new worker, she views that worker's $\boldsymbol{\Theta}_i$ and $\mathbf{L}_i$ as independent and identically distributed (iid) samples from the population distributions. At time $t = 0$, all potential workers, $i \in \mathscr{S}_0$, have the same history, $\mathscr{H}_{i,0} = \varnothing$; the same prior distribution for $\boldsymbol{\Theta}_i$, $\nu_{i,0} \equiv \hat{\nu}$; and no prior experience so that $n_{i,t} \equiv 0$. Thus, at time $t = 0$, the employer is indifferent among her choices. At any time $t > 0$, each worker, i, has cumulative experience $n_{i,t}$, and the employer uses i's employment history, $\mathscr{H}_{i,t}$, to update her beliefs concerning the distribution of the parameter $\boldsymbol{\Theta}_i$. We denote the posterior distribution that describes the employer's uncertainty concerning $\boldsymbol{\Theta}_i$ at time t as $\nu_{i,t}(X) = \mathbb{P}(\boldsymbol{\Theta}_i \in X \mid \mathscr{H}_{i,t})$, where $X \subseteq \Omega$ is any Borel set. For $\boldsymbol{\Theta}_i \sim \nu_{i,t}$ we let $Z_{i,t} \equiv Z(\nu_{i,t}, n_{i,t})$, and for $\{\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i\}$ we assume that worker i's performance $\{Z(\nu_{i,t}, n_{i,t}) \mid \boldsymbol{\theta}_i\}$ has density $\xi_{n_{i,t}}(z \mid \boldsymbol{\theta}_i)$. If worker i is employed at time t, then his experience, $n_{i,t}$, increases deterministically by one, and $n_{i,t+1} = n_{i,t} + 1$. Moreover, the employer updates her belief concerning i's ability distribution according to Bayes' rule. If $\mathscr{P}(\Omega)$ is the set of all probability measures, ν, on Ω, then the Bayes operator $\beta\colon \mathscr{P}(\Omega) \times \mathbb{R} \to \mathscr{P}(\Omega)$ is defined as

$$\beta(\nu_{i,t}, z)(X) = \frac{\int_X \xi_{n_{i,t}}(z \mid \boldsymbol{\theta})\, d\nu_{i,t}}{\int_\Omega \xi_{n_{i,t}}(z \mid \boldsymbol{\theta})\, d\nu_{i,t}} = \nu_{i,t+1}(X) \tag{5}$$

for each Borel subset $X \subseteq \Omega$. Thus, for any given observation, z, the Bayes operator maps the prior distribution, $\nu_{i,t}$, to its posterior distribution, $\nu_{i,t+1}$.

Within each period, t, the employer incurs a task-related cost that is driven by the selected employee's performance, $c(z_{i,t})$. We assume that $c(z)$ is continuous and nondecreasing in z, which reflects an efficiency-based measure of employee performance. Because the employer does not know employees' true abilities, in each period she uses her belief concerning the distribution of the current employee's ability, $\nu_{i,t}$, to estimate his expected task-related cost:

$$\mathbb{E}[c(Z(\nu_{i,t}, n_{i,t}))] = \int_\Omega \left(\int_{\mathscr{E}} c(g(\boldsymbol{\theta}, n_{i,t}, x)) \cdot \xi_{n_{i,t}}(g(\boldsymbol{\theta}, n_{i,t}, x) \mid \boldsymbol{\theta})\, dx \right) d\nu_{i,t}. \tag{6}$$

The employer also incurs costs that are specific to the hiring and retention policy she is implementing. If, at the start of a period, the employer hires a new employee, she incurs an *initial hiring* (or training) cost, c_h. If, at the end of a period, the employee quits, the employer bears a *quitting* cost, c_q, that includes potential separation costs and the cost of recruiting a replacement. If the employee does not quit, then the employer may decide to terminate him and switch to a different worker, in which case she bears a *switching* cost, c_s. Training, switching and quitting costs are assumed to be nonnegative. To properly account for switching and quitting costs, we introduce for each worker i and each time t a switching indicator, $u_{i,t}$, such that if policy π employs worker i over several, disjoint, time periods, then the indicator $u_{i,t}$ switches between 0 and 1, and it equals 1 at every time t such that worker i was not employed at $t-1$. Formally, we set $u_{i,0} = 1$ for all $i \in \mathscr{S}_0$ and for $t \geq 1$ we let

$$u_{i,t} = \begin{cases} 0 & \text{if } \pi(t-1) = i, \\ 1 & \text{if } \pi(t-1) \neq i. \end{cases}$$

When $\{i \in \mathscr{S}_0\colon u_{i,t} = 0\} \neq \{i \in \mathscr{S}_0\colon u_{i,t+1} = 0\}$, the workers employed at time $t-1$ and at time t differ, and the employer needs to incur the switching or quitting cost for the worker that was employed at time $t-1$.

For any time $\tau \ge 0$ and any state of prior distributions, experiences and switching indicators, $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = \{(\nu_{i,\tau}, n_{i,\tau}, u_{i,\tau})\colon\ i \in \mathcal{S}_0\}$, the infinite-horizon total expected discounted cost of any hiring and retention policy, π, from time τ onward is

$$\begin{aligned} C_\pi^\tau(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) &= \mathbb{E}\Bigg[\sum_{t=\tau}^{\infty} \gamma^t \big\{c_h \mathbb{1}(n_{\pi(t),t} = 0) + c(Z(\nu_{\pi(t),t}, n_{\pi(t),t})) \\ &\qquad + c_s u_{\pi(t),t} \mathbb{1}(\pi(t-1) \in \mathcal{S}_t \cap t > 0) \\ &\qquad + c_q u_{\pi(t),t} \mathbb{1}(\pi(t-1) \notin \mathcal{S}_t \cap t > 0)\big\}\Bigg], \end{aligned} \tag{7}$$

where the discount factor is $\gamma \in [0, 1)$. We note that in each period, t, the employer bears four possible sources of cost. The first, $c_h \mathbb{1}(n_{\pi(t),t} = 0)$, is the hiring and training cost for a new worker, and it is incurred only once, at the beginning of employee $\pi(t)$'s tenure. The second, $c(Z(\nu_{\pi(t),t}, n_{\pi(t),t}))$, reflects employee $\pi(t)$'s task-related costs. The third, $c_s u_{\pi(t),t} \cdot \mathbb{1}(\pi(t-1) \in \mathcal{S}_t \cap t > 0)$, is the cost of switching to a different worker at time t, should the previous employee be terminated. The fourth source of cost, $c_q u_{\pi(t),t} \cdot \mathbb{1}(\pi(t-1) \notin \mathcal{S}_t \cap t > 0)$, reflects the cost of switching to a different worker at time t, should the previous employee quit. When $t = 0$, no switching or quitting costs should be incurred, and we account for this by including the requirement $t > 0$ in the indicator functions in Equation (7). By observing that $\mathbb{1}(\pi(t-1) \in \mathcal{S}_t \cap t > 0) + \mathbb{1}(\pi(t-1) \notin \mathcal{S}_t \cap t > 0) = \mathbb{1}(t > 0)$, we rewrite (7) as

$$\begin{aligned} C_\pi^\tau(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) &= -c_s \mathbb{1}(\tau = 0) + \mathbb{E}\Bigg[\sum_{t=\tau}^{\infty} \gamma^t \big\{c_h \mathbb{1}(n_{\pi(t),t} = 0) \\ &\qquad + c(Z(\nu_{\pi(t),t}, n_{\pi(t),t})) + c_s u_{\pi(t),t} \\ &\qquad + (c_q - c_s) u_{\pi(t),t} \mathbb{1}(\pi(t-1) \notin \mathcal{S}_t \cap t > 0)\big\}\Bigg]. \end{aligned} \tag{8}$$

In this new formulation, the switching cost, c_s, is incurred any time the worker employed at time t is different from that employed at time $t-1$. The difference, $c_q - c_s$, then adjusts the value of the switching cost if the worker employed at $t-1$ has quit. The quantity, $-c_s \mathbb{1}(\tau = 0)$, outside the expectation compensates for the switching cost incurred for the first worker ever employed because $u_{i,0} = 1$ for all $i \in \mathcal{S}_0$.

We let Π denote the set of *nonanticipating* hiring policies, and we assume that the employer seeks a policy $\pi^* \in \Pi$ that minimizes the expected discounted value of future employment costs:

$$\pi^* \in \arg\min_{\pi \in \Pi} C_\pi^0(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}). \tag{9}$$

For the problem to be analytically tractable, we assume that the parameter space Ω is a Borel subset of $\mathbb{R}^d$, and we require that the single-period, task-related costs are uniformly bounded; that is, $c(g(\boldsymbol{\theta}, n, x)) \in [K_{\inf}, K_{\sup}]$ for each triple $(\boldsymbol{\theta}, n, x) \in \Omega \times \mathbb{N} \times \mathcal{E}$ (see, e.g., Sundaram 2005).

3. Structure of the Optimal Policy

The hiring and retention problem can be formulated as a Bayesian bandit problem with an infinite number of arms. Two elements of the problem complicate the analysis, however. First, when an employee quits, the arm associated with him becomes unavailable. Second, when the employer switches from one employee to another, she incurs the switching costs, c_s, that cannot be attributed to a single employee. In characterizing the optimal hiring and retention policy, we must address both of these difficulties.

3.1. Transformation to Problem with No Quitting

The fact that employees quit can be compensated for by transforming the problem with quitting to one in which workers are always available. Rather than quitting, they become *unproductive*, and their cost exceeds that of any productive worker. To do so, we assume that each employee, $i \in \mathcal{S}_0$, becomes unproductive at time t, after his $(n_{i,t} + 1)$st performance with probability equal to $q_{i, n_{i,t}}$ in (4). When employee i becomes unproductive at time t, his ability distribution changes from $\nu_{i,t}$ to $\nu_{i,t+1} = \mathbb{1}_K$, where $K \in (K_{\sup} + c_h + \max\{c_q, c_s\}, \infty)$ and $c(Z(\mathbb{1}_K, n)) = K$ for every n. Once employee i has become unproductive, he will never be able to go back to the productive state. The choice $K_{\sup} + c_h + \max\{c_q, c_s\} < K$ implies that the cost of an unproductive worker exceeds the cost of any possible realization of any productive worker, plus the largest cost of hiring a new worker. We then define the stopping time

$$\Lambda_i = \inf\{n_{i,t} \ge 1\colon\ c(Z(\nu_{i,t}, n_{i,t})) = K\} \tag{10}$$

as the time at which employee i becomes unproductive. Because an unproductive worker cannot go back to the productive state, we set $q_{i,k} = 0$ for all $k > n$ when $\Lambda_i = n$, and we modify the Bayes operator (5) as follows:

$$\beta(\nu_{i,t}, z)(X) = \begin{cases} \mathbb{1}_K & \text{if } \nu_{i,t} = \mathbb{1}_K, \\[2ex] \dfrac{\int_X \xi_{n_{i,t}}(z \mid \boldsymbol{\theta})\, d\nu_{i,t}}{\int_\Omega \xi_{n_{i,t}}(z \mid \boldsymbol{\theta})\, d\nu_{i,t}} & \text{otherwise.} \end{cases} \tag{11}$$

Call the original problem in (9), in which employees quit, Problem 1, and call the modified problem, in which they become unproductive, Problem 2. The following lemma confirms the fact that the problem with workers who become unproductive is analogous to that of those who quit.

LEMMA 1. (i) *In Problem* 2, *any policy that employs unproductive workers is never optimal.*

(ii) *A policy is optimal for Problem* 1 *if and only if it is optimal for Problem* 2.

Proofs of these claims and of the others below are found in the appendix.

Lemma 1 tells us that, for each policy $\pi \in \Pi$, employee i's working lifetime $\Lambda_i(\pi)$ in (3) and the time at which employee i becomes unproductive (10) are closely related. In fact, if employee i quits before he is terminated, then $1 + \Lambda_i(\pi) = \Lambda_i$. Otherwise, $1 + \Lambda_i(\pi) < \Lambda_i$.

LEMMA 2. *If* $\mathbb{E}[\Lambda_i] < \infty$, *then any policy for Problem* 1 *uses an infinite number of workers, almost surely* (*a.s.*).

Thus, if each employee's expected lifetime is finite, then the employer will end up hiring an infinite stream of employees in Problem 1. Similarly, an employer who avoids using employees who have become unproductive in Problem 2 will also use an infinite number of employees if $\mathbb{E}[\Lambda_i] < \infty$.

3.2. Transformation to Problem with Retirement Option

We derive the optimal policy for Problem 2 by solving a family of stopping problems in which, at each period, n, the employer chooses between employing a single worker, $i \in \mathcal{S}_0$, or terminating all employment and paying a so-called retirement cost, m. Given that we are considering an optimal stopping problem for a single employee, we drop the employee index, i, and the time index, t, from subscripts.

This approach, called the *retirement-option problem*, was introduced by Whittle (1980) for bandit problems with a finite number of arms and extended by Banks and Sundaram (1992) and Sundaram (2005) to study infinite-armed bandit models. In our context, the employer's problem is an infinite-horizon, discounted Markov decision process with uniformly bounded costs, a fact that implies that there exists an optimal hiring and retention policy that is stationary and deterministic (Bertsekas and Shreve 1978, Proposition 9.8).[1] The optimal value function for the retirement-option approach satisfies the following Bellman equation:

$$V(\nu, n, u, m) = \min\{m, HV(\nu, n, u, m)\}, \tag{12}$$

where

$$\begin{aligned} &HV(\nu, n, u, m) \\ &\quad = c_s u + c_h \mathbb{1}(n=0) + \mathbb{E}[c(Z(\nu, n))] \\ &\qquad + \gamma(1-q_n)\,\mathbb{E}[V(\beta(\nu, Z(\nu, n)), n+1, 0, m)] \\ &\qquad + \gamma q_n[c_q - c_s + V(\mathbb{1}_K, n+1, 0, m)]. \end{aligned} \tag{13}$$

[1] A policy is *stationary* if, at any time t, the action it prescribes depends only on the current state. A policy is *deterministic* if the action it prescribes is never randomized.

In words, at any decision time, the employer has the choice of retiring at cost m or continuing the employment of the worker currently on trial. The expected discounted cost of continuing, $HV(\nu, n, u, m)$, can be interpreted by looking at whether the employee is productive ($\nu \neq \mathbb{1}_K$) or not ($\nu = \mathbb{1}_K$). If the employee is productive, then with probability $1 - q_n$, he remains productive and $\beta(\nu, Z(\nu, n))$ is given by the bottom equation of (11) for each $Z(\nu, n) = z$, after dropping the i, t subscripts. With probability q_n, he becomes unproductive and his ability distribution changes to $\mathbb{1}_K$. If the employee is already unproductive at n, then $q_n = 0$, and the modified definition of the Bayes operator (11) gives us $\beta(\mathbb{1}_K, Z(\mathbb{1}_K, n)) = \mathbb{1}_K$. Here, we restrict our attention to values of m such that $m \le K/(1-\gamma)$, so that retiring is attractive when $\nu = \mathbb{1}_K$. Then, (13) becomes

$$\begin{aligned} &HV(\nu, n, u, m) \\ &\quad = c_s u + c_h \mathbb{1}(n=0) + \mathbb{E}[c(Z(\nu, n))] \\ &\qquad + \gamma(1-q_n)\,\mathbb{E}[V(\beta(\nu, Z(\nu, n)), n+1, 0, m)] \\ &\qquad + \gamma q_n[c_q - c_s + m]. \end{aligned} \tag{14}$$

If $\nu \neq \mathbb{1}_K$ and the employee is productive at n, the last addend represents the cost difference paid for an employee who has quit, $c_q - c_s$, plus the retirement cost for the employer, m. The quantity $HV(\nu, n, u, m)$ hence represents the cost of employing a worker with ability distribution, ν, experience, n, and switching indicator, u, for at least one period, followed by an optimal termination decision that depends on the retirement payment, m.

The stopping time

$$\begin{aligned} &\tilde{\Lambda}(\nu, n, u, m) \\ &\quad = \inf\{r \ge 1\colon HV(\nu_r, n+r, u_r, m) > m\} \end{aligned} \tag{15}$$

is the time at which the employer chooses to retire, and $\{\nu_r\}_{r\ge1}$ and $\{u_r\}_{r\ge1}$ represent the evolution of the ability distribution and the switching indicator after period n. For $r = 0$, we set $\nu_0 \equiv \nu$ and $u_0 \equiv u$.

Let $Q_n = \{\omega\colon \Lambda > n, \tilde{\Lambda}(\nu, n, u, m) = \Lambda - n\}$ be the set of sample paths for which a productive worker with ability distribution, ν, experience, n, and switching indicator, u, quits before he is terminated. Note that if a worker is already unproductive at n and $\nu = \mathbb{1}_K$, then $\Lambda \le n$, and therefore $Q_n = \varnothing$. Then, we can write the expected discounted cost of continuing (14) as

$$\begin{aligned} &HV(\nu, n, u, m) \\ &\quad = \mathbb{E}\Bigg[c_s u + c_h \mathbb{1}(n=0) + \sum_{r=0}^{\tilde{\Lambda}(\nu, n, u, m)-1} \gamma^r c(Z(\nu_r, n+r)) \\ &\qquad + \gamma^{\tilde{\Lambda}(\nu, n, u, m)}\{(c_q - c_s)\mathbb{1}_{Q_n} + m\}\Bigg]; \end{aligned} \tag{16}$$

this last representation and its properties will be crucial in the proofs of many of our results.

Given the availability of the value function (12), we are interested in the value of m for which the employer is indifferent between continuing to employ the current hire or retiring, at cost m. We denote that value by the index

$$M(\nu, n, u) = \sup\{m \in \mathbb{R}\colon V(\nu, n, u, m) = m\}. \tag{17}$$

This index is well defined because the value function (12) is concave and nondecreasing in m, a fact that is stated and proved in the appendix. It is a direct analogue of the definition of the Gittins index proposed by Whittle (1980) for problems without learning or switching costs. Asawa and Teneketzis (1996) propose a corresponding index for problems with switching costs.

3.3. Optimal Policy

When the employer switches from one employee to another, she incurs a switching cost, a fact that can make the characterization of optimal policies difficult. In particular, when the set of available hires is finite, an employer that switches away from and then returns to an employee, i, at a later period pays a switching cost that she would not have incurred had she continued to employ i over contiguous periods (Banks and Sundaram 1994).

A number of researchers have sought to characterize optimal policies for such bandit problems with switching costs. For problems with a finite number of arms, Asawa and Teneketzis (1996) define two indices, a traditional Gittins index analogous to (17), along with a corresponding "switching cost index," and they show that these indices can be used to describe necessary, though not sufficient, conditions under which an optimal policy will switch arms. Niño-Mora (2008) shows how to efficiently calculate Asawa and Teneketzis's (1996) indices. As a part of his analysis, Niño-Mora (2008) shows that (for finite arms), if it is not optimal at t to use an arm that was used at $t-1$, then it would also not be optimal to use that arm at t had the arm *not* been used at $t-1$. In the context of problems with infinite sets of arms, Bergemann and Välimäki (2001) independently make a similar observation. Bergemann and Välimäki (2001) further note that, in problems with an infinite set of a priori identical, "untried" arms once it is optimal to switch away from an arm, i, to use another that has not been tried, it will never pay to switch back to i, since there will always remain another untried arm that will be preferable to i.

Bergemann and Välimäki (2001) use the "forward induction" formulation of Gittins (1979) to prove these results for infinite-armed Bayesian bandits that do not evolve with experience and without switching costs, and they then sketch an argument for extending the results to problems with switching costs. For arms without switching costs that evolve, the proof in the appendix is based on the retirement option formulation in Whittle (1980) and follows the line of reasoning in Sundaram (2005, Appendix A). We then apply the argument sketched in Bergemann and Välimäki (2001) to explicitly prove the extension to problems with switching costs. The following proposition summarizes the main results.

Proposition 1 (Follows Sundaram 2005 and Bergemann and Välimäki 2001). *Let $\nu_{i,0} = \hat{\nu}$, $n_{i,0} = 0$, and $u_{i,0} = 1$ for all $i \in \mathcal{S}_0$, and let $\hat{m} = M(\hat{\nu}, 0, 1)$ be the index of a worker who has not yet been tried.*

(i) *A policy π^* is optimal if and only if*

$$\pi^*(t) \in \left\{ i \in \mathcal{S}_0\colon M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) = \inf_{j \in \mathcal{S}_0} M_j(\nu_{j,t}, n_{j,t}, u_{j,t}) \right\}, \quad \text{a.s. for all } t = 0, 1, 2, \ldots.$$

(ii) *At any time, t, at most one worker, i, has Gittins index $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) < \hat{m}$.*

(iii) *Let $t_i = \inf\{t\colon \pi^*(t) = i\}$ be the first time worker i is employed. Under the optimal policy π^* in* (i):

(a) *Worker i is employed continuously for $\Lambda_i(\pi^*)$ periods: $\pi^*(t) = i$ for all $t_i \le t < t_i + \Lambda_i(\pi^*)$.*

(b) *It is never optimal to employ worker i from time $t_i + \Lambda_i(\pi^*)$ on: $\pi^*(t) \neq i$ for all $t \ge t_i + \Lambda_i(\pi^*)$.*

(iv) *If $\mathbb{E}[\Lambda_1] < \infty$, then $\hat{m} - c_s = \inf_{\pi \in \Pi} C_\pi^0(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$.*

Given the structure of the optimal policy in part (i) of Proposition 1, we can justifiably call (17) a *Gittins index*. Moreover, when the optimal policy is implemented, part (ii) implies that there is often just one Gittins-index-minimal employee. Part (iii) shows that it is never optimal to employ a worker who was previously replaced. For an employer seeking to retain a single employee, the hiring and retention problem, therefore, decomposes into a sequence of iid optimal stopping problems: hire an employee from the pool and retain him until he turns over or his Gittins index rises above $\hat{m}$, whichever comes first. In turn, the optimal policy yields a discounted renewal reward process, with expected value described in part (iv).

Part (iv) of Proposition 1 directly links the expected total discounted cost of the optimal policy to the Gittins index, a result that does not generally hold in bandit problems with finite numbers of arms. In §5, we use the result to estimate the expected discounted value of a Gittins-index policy.

4. Extensions: Multiple Parallel Workers and Different Pools

Sections 2 and 3 consider the problem of employing a single worker. We now consider two extensions.

Section 4.1 considers the problem in which distinct (infinite) pools of heterogeneous workers are available. Section 4.2 considers an employer who wishes to retain multiple employees who work in parallel. In both cases, the optimality of an index rule is retained.

4.1. Heterogeneous Populations

When the employer faces a finite number of heterogeneous populations, her optimal hiring and retention policy is the same as the one proposed in Proposition 1, part (i). For example, consider two infinite pools $\mathcal{S}_0^{\nu}$ and $\mathcal{S}_0^{\eta}$, for which the untried workers have common prior distributions $\hat{\nu}$ and $\hat{\eta}$, with $\hat{\nu} \neq \hat{\eta}$. Let $M(\hat{\nu}, 0, 1)$ and $M(\hat{\eta}, 0, 1)$ be the indices of the untried workers in each pool. If $M(\hat{\nu}, 0, 1) \neq M(\hat{\eta}, 0, 1)$, then workers belonging to the pool with larger index are never employed by an optimal policy. Otherwise, if $M(\hat{\nu}, 0, 1) = M(\hat{\eta}, 0, 1)$, then the employer is indifferent between the two populations.

4.2. Hiring and Retention of Multiple Workers

Assume now that $\nu_{i,0} = \hat{\nu}$, $n_{i,0} = 0$, and $u_{i,0} = 1$ for all $i \in \mathcal{S}_0$, and consider the hiring and retention problem in which the employer wishes to retain a fixed number, D, of people working in parallel.

One can partition the infinite pool of potential employees, $\mathcal{S}_0$, into D separate, countably infinite pools, $\mathcal{S}_{1,0}, \ldots, \mathcal{S}_{D,0}$, of identical workers with common prior distribution, $\hat{\nu}$, no experience, and common switching indicator equal to 1. When employee i in pool d quits at time t, he is removed from that pool so that $\mathcal{S}_{d,t+1} = \mathcal{S}_{d,t} \backslash \{i\}$. Then, the infinite-horizon total expected discounted cost is

$$\begin{aligned} C_{\pi}^{0,D}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) &= \mathbb{E}\bigg[\sum_{t=0}^{\infty} \gamma^t \sum_{d=1}^{D} \big(c_h \mathbb{1}(n_{\pi_d(t),t} = 0) + c(Z(\nu_{\pi_d(t),t}, n_{\pi_d(t),t})) \\ &\quad + u_{\pi_d(t),t} \mathbb{1}(t > 0)\big\{c_s \mathbb{1}(\pi_d(t-1) \in \mathcal{S}_{d,t}) \\ &\quad + c_q \mathbb{1}(\pi_d(t-1) \notin \mathcal{S}_{d,t})\big\}\big)\bigg], \qquad (18) \end{aligned}$$

where $\pi_d(t) \in \mathcal{S}_{d,t}$ identifies the index of the worker who is employed from pool d at time t, $\nu_{\pi_d(t),t}$ his ability distribution, $n_{\pi_d(t),t}$ his experience, and $u_{\pi_d(t),t}$ his switching indicator value. By interchanging the sums in (18), one obtains $C_{\pi}^{0,D}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = \sum_{d=1}^{D} C_{\pi}^{0,d}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, where $C_{\pi}^{0,d}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ is the dth position's expected discounted cost, as defined in (7). Thus, the D positions' costs are separable so that the total expected discounted cost is minimized when a Gittins-index-minimal worker is employed in each pool.

At any time, t, at which the employer seeks to hire a new worker for any of the D positions, she can employ any untried worker who belongs to the pool of potential employees, $\mathcal{S}_t$. This result, which follows directly from Bergemann and Välimäki (2001), crucially depends on the assumption that all workers have the same experience and ability distribution at time $t = 0$, so that the artificial splitting of potential hires into D pools is possible.

We note that our analysis of multiple employees also hinges on the independence of the outcomes of various employees' tasks. In many settings, task outcomes may be correlated across workers, however, and the optimality of an allocation index is no longer valid, as for other bandit problems with correlated arms. One potentially promising avenue for addressing such correlations in future work is the knowledge gradient approach (Frazier et al. 2009).

5. Implementing the Optimal Policy

This section shows how analytic properties of the hiring and retention problem can be combined with dynamic programming to enable the computation of the relevant Gittins indices when performance has certain structural properties. As shown in the appendix, for any given ν, n, and u, the value function, $V(\nu, n, u, m)$, is concave and nondecreasing in m. Therefore, given ν, n, and u, a simple search scheme, such as a bisection, can be used to find the largest fixed point, $M(\nu, n, u)$, that defines the Gittins index.

Because our set of iid stopping problems allows us to focus on a single employee, we drop the indices i and t as subscripts and let $Z_n = g(\boldsymbol{\theta}, n, \epsilon_n)$. To calculate solution values, we explicitly define the functional form of the $(n+1)$st performance for a worker, Z_n. We assume that $g(\cdot)$ is invertible and that

$$g^{-1}(Z_n) = A + h(n) + \epsilon_n, \quad n = 0, 1, 2, \ldots \qquad (19)$$

is a linear model where A determines an unknown base level that may vary across workers, $h(n)$ is a known learning function, and ϵ_n is normally distributed noise with mean 0 and known variance σ^2.

Because A is unknown, the mean of the noise can be assumed to be zero without loss of generality. We assume that the potential hire's base level of performance, A, has initial prior distribution, $\hat{\nu}$, that is normally distributed with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$, $N(\hat{\mu}, \hat{\sigma}^2)$. The form in (19) implies another structural property that will be useful for computing the Gittins indices of workers. The random variables $g^{-1}(Z_n) - h(n)$ are normally distributed with unknown mean A and variance $\sigma^2 + \hat{\sigma}^2$. By standard Bayesian analysis, ν, the posterior distribution of A after observing n tasks, $\mathbf{z}_n = (z_0, z_2, \ldots, z_{n-1})$, is normal with

$$\mathbb{E}[A \mid \mathbf{z}_n] = \frac{\hat{\mu}(\sigma^2/\hat{\sigma}^2) + \sum_{k=0}^{n-1}(g^{-1}(z_k) - h(k))}{n + \sigma^2/\hat{\sigma}^2} \quad \text{and}$$

$$\operatorname{Var}[A \mid \mathbf{z}_n] = \frac{\hat{\sigma}^2\sigma^2}{\sigma^2 + n\hat{\sigma}^2}.$$

Define $\hat{p} = \sigma^2/\hat{\sigma}^2$, and let $p = \hat{p} + n$, where n is the number of samples observed for the single-worker problem. Set $y_p = \hat{\mu}\hat{p} + \sum_{k=0}^{n-1}(g^{-1}(z_k) - h(k))$ and $w_p = y_p/p$. The posterior distribution, ν, of A given $\mathbf{z}_n$ is thus $N(w_p, \sigma^2/p)$. We can therefore describe (ν, n) by (w_p, p).

These assumptions are sufficient to guarantee that both the Bellman equation (12) and the Gittins index (17) are monotone in the posterior mean of A, w_p.

Proposition 2. *For any given p, u, and m, the value function $V(w_p, p, u, m)$ is nondecreasing in w_p. For any given p and u, the Gittins index $M(w_p, p, u)$ is nondecreasing in w_p.*

The monotonicity of the Gittins index with respect to w_p allows us to concisely describe the optimal policy. For each $p = \hat{p} + n$, there is a simple "stopping" boundary, $\mathfrak{b}(p)$, such that it is optimal to retain the employee (continue) if $w_p < \mathfrak{b}(p)$ and to terminate the employee (stop) if $w_p > \mathfrak{b}(p)$.

Arlotto et al. (2010) provide more detail for how to use the above results to approximate V and the stopping boundary, $\mathfrak{b}$, when (19) applies; the functions g and h are known and finite for finite values of their arguments; the noise, ϵ_n, has zero mean and known sampling variance, σ^2; and the prior distribution for A is $N(w_{\hat{p}}, \sigma^2/\hat{p})$, so that Proposition 2 applies. In summary, we use the common technique of approximating the evolution of the posterior distribution as samples are observed, a Gaussian process, with the evolution of the posterior distribution of a related trinomial process on a grid. We construct the necessary grid of points in the (w, p) coordinate system, estimate the terminal conditions (the period at which the dynamic programming backward recursion starts, typically a large number of periods in the future) using Monte Carlo simulation, perform a backward recursion using a trinomial tree approximation on the grid of points to approximate both V and the optimal stopping boundary for a given value of m, and then search for the value of m that identifies the Gittins index. This process also identifies the optimal stopping boundary that determines the optimal solution to the hiring and retention problem.

The numerical results in §6 correspond to a learning function that sets $g(z) = e^z$ and $h(n) = b\ln(n+1)$. This corresponds to (2) with a common learning parameter $b_i = b$ and

$$\ln(Z_n) = A + b\ln(n+1) + \epsilon_n, \quad n = 0, 1, 2, \ldots, \tag{20}$$

where $\epsilon_n \sim N(0, \sigma^2)$. Here, (20) is consistent with empirical studies of various industries. For example, Brown et al. (2005), Shen (2003), and Shen and Brown (2006) provide evidence that handle times for call-centers are frequently lognormally distributed.

The above approach can be used to numerically evaluate other forms of $h(\cdot)$, and we have also tested $h(n) = b\ln(1 + n/(n + \zeta_1))$ and $h(n) = b\ln(1 + \min\{n, \zeta_2\})$. Although the details of the stopping boundaries can change with the functional form, the qualitative conclusions we reach from numerical tests with these functions are analogous to what we describe in §6. Similarly, we can define a as a common, known parameter and $g^{-1}(Z_n) = a + Bh(n) + \epsilon_n$ to model pools of workers with a common base level of quality and heterogeneous rate of learning. Although the theoretical results described in §3 hold for even more complex settings, such as those with heterogeneous and unknown A and B, the numerical approach here becomes more difficult. In particular, stopping boundaries become multidimensional, and monotonicity results, such as those described in Proposition 2, may not hold.

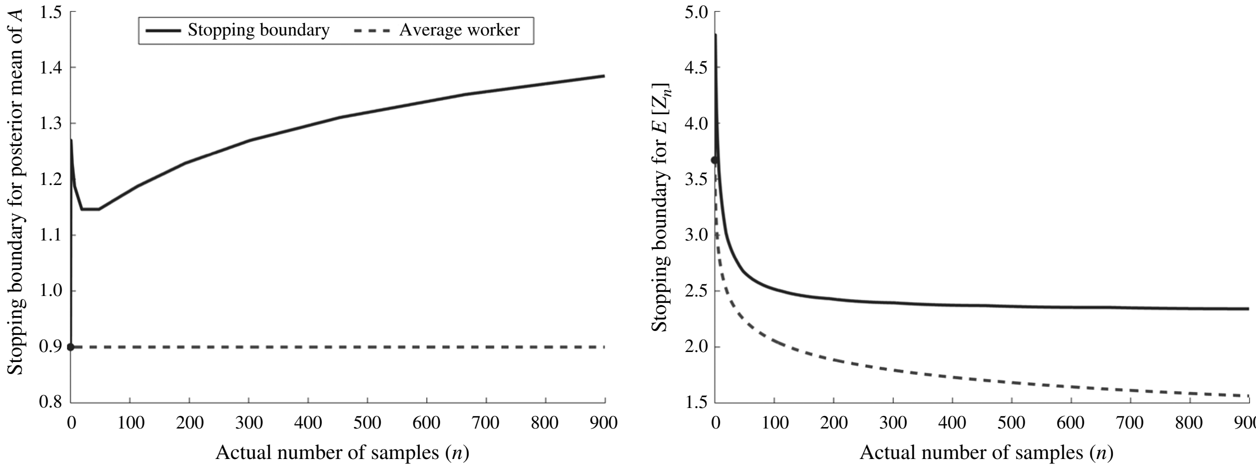
6. Numerical Examples and the Value of Screening

In this section, we use the methods described in §5 to calculate Gittins indices, as well as associated optimal stopping boundaries, for several examples. We also use discrete-event simulation to estimate rates of termination and voluntary turnover. We compare the performance of the optimal Gittins-index policy with that of other easily implementable policies, and we demonstrate that an active hiring and retention policy reduces costs and improves the pool of workers who are employed. We perform a sensitivity analysis with respect to the key parameters of our model, and we conclude that increases in employee learning rates reduce costs, improve the pool of employed workers, and lower termination rates. Moreover, we observe that managers favor pools of potential workers with a broader set of abilities.

6.1. Balancing Uncertainty and Learning Effects

The first example is loosely motivated by a call center. Each Z_n represents the average duration (in minutes) of the calls that an agent handles after n days of experience. We use the log-linear learning curve model (20). The distribution of the base-level performance parameter, A, has mean $\hat{\mu} = 0.90$ and standard deviation $\hat{\sigma} = 0.40$, and the sampling standard deviation in the daily average of the service times is $\sigma = 0.80$. This implies an expected service time of untried agents of $\mathbb{E}[Z_0] = 3.67$. The annual discount rate is 10%, so the one-period discount rate is $\gamma = 0.9996$ (based on a year of 250 days), and the cost function is linear, $c(z) = cz$, with unit cost $c = 1$. The training cost is $c_h = 30$, which corresponds to the expected cost of employing untried workers for approximately 10 days (two weeks). Termination and quitting costs are set

Figure 1 Stopping Boundaries for Posterior Mean of A (Left) and for $\mathbb{E}[Z_n]$ (Right)



Note. Parameters: $b = -0.1255$, $\hat{\mu} = 0.90$, $\hat{\sigma} = 0.40$, $\sigma = 0.80$, $\hat{s} = 4$, $c_h = 30$, $c_s = c_q = 0$.

equal to 0. (See Theorem 1 in §6.3.2.) Learning is deterministic with rate $b = \ln \alpha / \ln 250$, where $\alpha \in (0, 1]$ represents the amount of learning accrued in the first year of tenure so that $\mathbb{E}[Z_{249}] = \alpha \mathbb{E}[Z_0]$. Choosing $\alpha = 0.50$, we obtain $b = -0.1255$.

For lack of real-world data concerning turnover behavior, and to focus our numerical results on the effects of learning, we assume that the quitting probability q_n is constant over time. We let $q_n = 0.01$ for all n, so (in the absence of termination) workers turn over, on average, every 100 days.

Figure 1 displays the stopping boundary associated with the Gittins index for untried employees who, in this example, have $\hat{m} = 5{,}491.7$. The left panel plots the stopping boundary with respect to the posterior mean of A, and the solid line in the right panel plots the analogous stopping boundary with respect to the posterior mean of Z_n. From Proposition 2, we know that an employee whose posterior means fall below these stopping boundaries has a Gittins index below $\hat{m}$ and should be retained, and one whose posterior means fall above the stopping boundary should be replaced by a new hire.

In the left panel, we see that the stopping boundary with respect to the posterior mean of A has an interesting shape. The initial jump from the prior mean, $\hat{\mu} = 0.90$, up to 1.27 is attributed to the elimination of the training cost, c_h, which is incurred only on day 0. Afterward, the stopping boundary has a "cupped" shape for the first few periods of an employee's tenure. The dip reflects the effect of statistical learning on the part of the employer. As more samples are collected, uncertainty about the "true" quality of the worker decreases, and the employer can screen workers on the basis of a more informative prior distribution. The subsequent climb reflects the gains the employee enjoys as on-the-job experience makes even relatively poor-quality workers attractive candidates for retention. In its rightmost reaches, the curve appears to increase to an asymptote involving a constant minus $h(n)$ (a phenomenon that was observed for other learning functions we tested).

The right panel shows the stopping boundary with respect to $\mathbb{E}[Z_n]$. Here, the stopping boundary is unimodal, with a peak on day 1 due to the elimination of the day 0 training cost, followed by a monotone decrease that is initially steep and that later flattens out. Unlike the left panel, the right panel does not explicitly display a "dip" that reflects the problem's two conflicting forces, between the employer's statistical learning and the employees' learning by doing. Instead, after day 1, we find a monotonically decreasing stopping boundary that requires a worker's expected performance to keep improving over time. The dashed line in both panels plots the prior mean, $\hat{\mu}$ (left), and the expected call times, $\mathbb{E}[Z_n]$ (right), for an "average" employee with base-level service time $A = \hat{\mu}$. The vertical distance between the two curves is a measure of how much better or worse a "marginally retained" employee is in comparison wih an "average" employee. The presence of training costs induces managers to retain workers who are worse than average.

The simulation results in Table 1 describe how the optimal policy affects employee retention. The results are based on 50,000 trials of the single-worker optimal stopping problem, and they show the fraction of workers who are terminated or quit within various time windows.

The policy terminates 39.82% of the employees: 1.96% of workers are terminated on day 1, 28.30% are terminated during periods 2–10, and 9.57% thereafter. Hence, much of the termination occurs early on. Of course, termination rates vary significantly with

Table 1 **Optimal Policy and Employee Retention**

	Day 1	Days 2–10	Days 11–20	Total
Terminated workers	0.0196 (0.0006)	0.2830 (0.0020)	0.0557 (0.0010)	0.3982 (0.0022)
Workers who quit	0.0102 (0.0005)	0.0692 (0.0011)	0.0539 (0.0010)	0.6018 (0.0022)

Note. Standard errors for the mean in parentheses.

training costs. In §6.3 we present a sensitivity analysis that addresses this relationship.

6.2. How the Optimal Policy Compares with Simpler Policies

This section compares the optimal policy with four families of alternative hiring policies. In the first family, workers are never terminated, and they serve until they naturally turn over. In the second, workers are monitored for a limited screening period, during which they can be terminated after each day of performance. If retained at the end of the screening period, they are never terminated. In Table 2 we report results for this type of policy when the screening period is 5, 10, or 20 days long. The third family considers Gittins-index policies in which workers are screened and termination can occur every 5, 10, or 20 days of performance. (Note that the optimal policy described in this paper is a Gittins-index policy in which screening takes place each day.) Finally, the fourth family considers policies with a trial period of a given length (1, 5, 10, or 20 days) within which workers are not terminated. At the end of the trial period, the employer decides whether to retain or terminate the worker, and, if he is retained, he is not terminated until he turns over. In all cases, we use optimal retain/terminate thresholds, given the details of the particular policy.

Table 2 reports infinite-horizon total expected discounted costs, termination rates, long-run average service rates, and the expected discounted number of monitored periods for each policy. The results reported are obtained by simulating 1,000 trials with enough workers to cover 50,000 time periods within each trial. We also report analogous simulation results for the optimal policy and note that, because it is estimated via simulation rather than backward recursion, the Gittins index for this example varies slightly (within one standard error) from that reported in §6.1.

The results in the second column of Table 2 show that the optimal policy we examined leads to a substantial reduction in cost. For instance, the policy that does not screen employees has a total expected discounted cost that is 10.41% higher than that of the optimal Gittins-index policy. We already know from Table 1 that most termination in the optimal policy occurs relatively early in employees' tenure. It is not surprising, then, that the policy that screens workers in each of the first 20 days performs nearly as well as the optimal one. Interestingly, the Gittins-index policy that screens workers every five days also performs close to optimally. Thus, screening needs not to occur every period for a policy to be effective. The results for "one-shot" at 5 and 10 periods also suggest that simple, one-shot retention decisions have the potential to perform well, with average discounted costs within a few percent of the optimal Gittins-index policy.

For any hiring policy, π, its long-run average service rate is

$$\mu(\pi)^{-1} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\mathbb{E}[Z_{\pi(t),t}]},$$

Table 2 **Comparison with Other Hiring Policies**

Policy	Total expected discounted cost		Fraction of terminated workers	Long-run average service rate		Expected discounted number of monitored periods
Optimal policy	5,494.1 (12.3)	—	0.3948 (0.0005)	0.6417 (0.0138)	—	2,333 (6.6)
Never screen	6,066.3 (15.4)	10.41%	0.0000 (0.0000)	0.5364 (0.0149)	−16.41%	0 (0.0)
Screen 1–5	5,618.3 (12.4)	2.26%	0.3540 (0.0006)	0.6179 (0.0140)	−3.71%	149 (21.3)
Screen 1–10	5,539.8 (11.9)	0.83%	0.3764 (0.0006)	0.6315 (0.0140)	−1.59%	288 (39.9)
Screen 1–20	5,505.3 (11.7)	0.20%	0.3960 (0.0005)	0.6405 (0.0137)	−0.19%	525 (67.5)
Gittins every 5	5,528.9 (11.9)	0.63%	0.3690 (0.0005)	0.6341 (0.0134)	−1.18%	445 (4.9)
Gittins every 10	5,569.3 (11.5)	1.37%	0.3282 (0.0005)	0.6218 (0.0136)	−3.10%	212 (4.2)
Gittins every 20	5,672.2 (12.4)	3.24%	0.2623 (0.0005)	0.6015 (0.0130)	−6.27%	97 (3.5)
One-shot at 1	5,896.0 (14.4)	7.32%	0.2432 (0.0005)	0.5739 (0.0153)	−10.57%	24 (3.4)
One-shot at 5	5,639.6 (12.7)	2.65%	0.3228 (0.0006)	0.6108 (0.0139)	−4.81%	23 (3.1)
One-shot at 10	5,644.4 (12.2)	2.73%	0.3234 (0.0006)	0.6146 (0.0137)	−4.23%	21 (2.7)
One-shot at 20	5,696.1 (12.3)	3.68%	0.2669 (0.0005)	0.6004 (0.0131)	−6.43%	18 (2.2)

Note. Standard errors for the mean in parentheses.

the long-run average number of calls that an agent handles per minute each day. Its numerical values are reported in fourth column of Table 2, and they suggest that the optimal Gittins-index policy leads to an overall improvement of employee performance. Moreover, the quantity $\mu(\pi)^{-1}$ can then be used to obtain a rough estimate of the number of agents needed for a given call volume. For instance, if we compare the optimal Gittins-index policy with the "never screen" policy, we see that the former requires, on average, 16.41% fewer workers to maintain the same level of capacity. To more clearly understand this, consider the hypothetical scenario in which a call center has an average load of 53.64 calls per minute. With the optimal policy, this requires employing $53.64/0.6417 = 83.59$ workers—long-run average—to have a "fully loaded" system. With the never screen policy, the same fully loaded system requires $53.64/0.5364 = 100$ workers, and the optimal policy employs 16.41% fewer workers.

The rightmost column of Table 2 counts the expected discounted number of periods in which the employer monitors the performance of its employees. Naturally, the optimal Gittins-index policy in which monitoring occurs every day is the most expensive along this dimension. Interestingly, the "screen 1–20" and "Gittins every 5" policies perform well with respect to costs and require approximately one-fourth of the monitoring effort on the part of the employer. Thus, to the extent that monitoring is an expensive activity, the nature of effective policies may change. While the explicit representation and optimization of monitoring is beyond the scope of the current paper, it certainly merits future work.

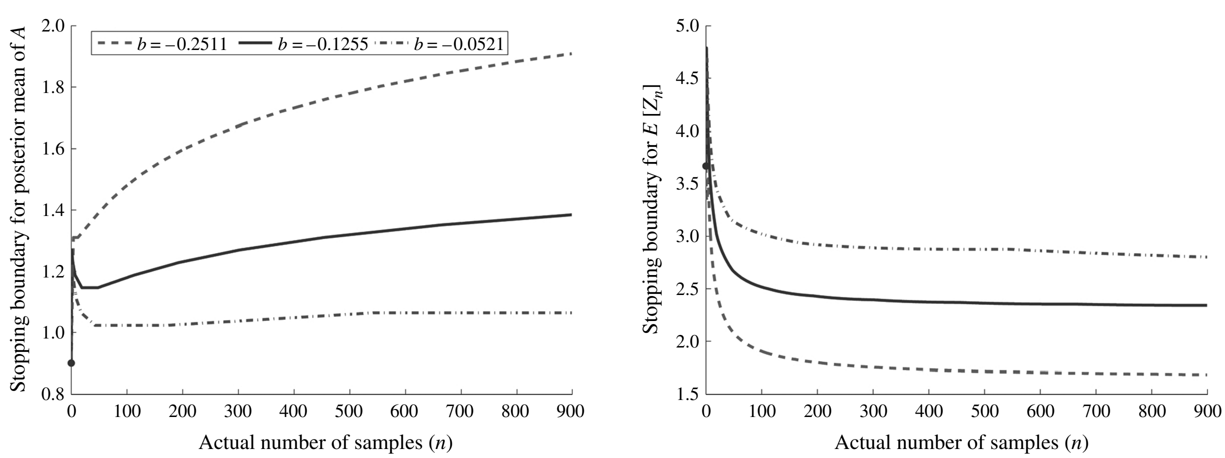
6.3. Sensitivity Analysis

This section examines how the optimal policy depends on key parameters: employees' learning rates; switching and quitting costs; employer uncertainty regarding employee performance; task-by-task variability; and training costs. The Gittins indices, turnover, and termination rates reported in this section are computed as in §6.1.

6.3.1. Learning Rates. Section 6.1 studies a pool of workers whose performance improves by 50% over the first 250-day year ($b = -0.1255$). Here, we compare this performance with that of fast-learning workers who improve by 75% in the first one year ($b = -0.2511$), as well as that of slow-learning workers who improve only by 25% in the same amount of time ($b = -0.0521$). All other parameters are as in §6.1.

Figure 2 plots the stopping boundary with respect to the posterior mean of A (left) and with respect to $\mathbb{E}[Z_n]$ (right) in these new settings. In the left panel, we notice that the "cupped" shape of the stopping boundary in the early stages of employment is more prominent for the slow learners, and the set of their allowable posterior means is smaller. On the other hand, the fast-learning workers immediately benefit from a tangible performance improvement in their first few days so that the "cupped" part of the stopping boundary disappears. The contribution of this experience-based learning is so high that the screening policy retains workers with a broader set of posterior means. With a faster learning rate, every employee is faster for each level of experience, and one expects the stopping boundary with respect to $\mathbb{E}[Z_n]$ to decline. This is indeed the case, and, in the right panel of Figure 2, we see that the stopping boundary for fast-learning workers is the bottom one. A similar argument explains why the stopping boundary for slow learners is the top one in the right panel. To more clearly understand the effect of changes in employees' learning, we also look at the

Figure 2 **Stopping Boundaries for Different Learning Rates**



Note. Other parameters: $\hat{\mu} = 0.90$, $\hat{\sigma} = 0.40$, $\sigma = 0.80$, $\hat{s} = 4$, $c_h = 30$, $c_s = c_q = 0$.

Table 3 Simulation Results with Different Learning Rates

b	Gittins index	Fraction of terminated workers				Long-run average service rate
		Day 1	Days 2–10	Days 11–20	Total	
−0.2511	3,905.6	0.0102 (0.0004)	0.1834 (0.0017)	0.0275 (0.0007)	0.2366 (0.0019)	1.0253 (0.0286)
−0.1255	5,491.7	0.0196 (0.0006)	0.2830 (0.0020)	0.0557 (0.0010)	0.3982 (0.0022)	0.6417 (0.0138)
−0.0521	6,762.1	0.0334 (0.0008)	0.3167 (0.0021)	0.0822 (0.0012)	0.4885 (0.0022)	0.4972 (0.0089)

Note. Standard errors for the mean in parentheses.

values of the Gittins index, at the fraction of terminated workers, and at the long-run average service rate for these three b's. Table 3 shows that the optimal retention policy for pools of fast learners generates the smallest infinite-horizon expected discounted cost, the lowest fraction of terminated workers, and the largest service rate. Conversely, slow learners are the most expensive, have the highest termination rates, and have the lowest long-run average service rates.

The results of Table 3 suggest a potentially important, positive sequence of managerial implications. Improvements in on-the-job learning rates make employees with relatively poor initial abilities more attractive compared with untried employees, and it is optimal for the employer to retain them. As a consequence, optimal termination rates decline. Thus, improvements in on-the-job learning rates may allow the employer to enjoy a secondary benefit of being able to retain a wider array of employees. Moreover, there is evidence from the management literature that lower rates of termination may make a company a more desirable place to work and improve its pool of potential hires (Huselid 1995). Such an employee response to changes in the employment policy is of potential interest. As noted in the introduction, explicit treatment of the phenomenon would extend our analysis into the realm of stochastic games, however.

Remark 1. Empirical evidence in the learning literature shows that slower learners can produce higher value in the long run (see, e.g., March 1991, Uzumeri and Nembhard 1998). In our model, this effect could be investigated by segmenting slow learners and fast learners in two different populations. If the prior ability distribution in each population were known, then the optimal policy would be as in §4.1, and only workers belonging to the population with better index would be employed. If the prior ability distributions were unknown, however, one would need to construct a hierarchical model that goes beyond the scopes of the current paper.

6.3.2. Switching and Quitting Costs. One would expect that changes in switching and quitting costs would similarly affect the optimal policy. However, the theorem below shows that when the quitting probabilities are constant—so that $q_{i,n} = q$ for all n and for all $i \in \mathcal{S}_0$—this is not the case.

To state the theorem we need to keep track of how the training, quitting, and switching costs affect the Gittins index. To that end, we modify our notation to account for these differences, letting $M(\nu, n, u, c_h, c_s, c_q)$ be the Gittins index (17) and $\widehat{m}(c_h, c_s, c_q) = M(\hat{\nu}, 0, 1, c_h, c_s, c_q)$.

Theorem 1. *Assume that $\nu_{i,0} = \hat{\nu}$, $n_{i,0} = 0$, and $u_{i,0} = 1$ for all $i \in \mathcal{S}_0$. Then, if the quitting probabilities are constant, i.e., $q_{i,n} = q$ for all $i \in \mathcal{S}_0$ and all n, $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}, c_h, c_s, c_q) < \widehat{m}(c_h, c_s, c_q)$ if and only if $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}, c_h + c_s, 0, 0) < \widehat{m}(c_h + c_s, 0, 0)$, for all $t \geq 0$.*

Thus, if the hazard rate for quitting is constant for all employees at all times, then changes in switching and quitting costs do not affect the relative ordering of workers' Gittins indices. Of course, the values of the Gittins indices change, as do the (analogous) expected discounted costs of the problem. However, because the relative orderings do not change, changes in the switching and quitting costs do not affect the optimal policy; therefore, we do not report a sensitivity analysis with respect to c_s or c_q.

When the quitting probabilities are *not* constant, the specifics of the optimal policy can change with c_s and c_q. Nevertheless, the overall structure of the optimal policy does not change. Proposition 1 holds for any quitting behavior $q_{i,n}$, as in (4).

6.3.3. Variance of the Base-Level Performance, Variance of Samples, and Training Costs. A sensitivity analysis for the prior distribution of abilities, the sampling variance, and training costs yields intuitive results for which we provide a brief overview:

- The Gittins index reflects an option value inherent in the ability to change arms and favors arms with more diffuse prior distributions. A sensitivity analysis for the variance of the prior distribution of abilities agrees with the general idea: for a given $\hat{\mu}$, an increase in the variance, $\hat{\sigma}^2$, of ability across workers allows the employer to screen more strictly, thereby increasing termination rates, retaining relatively more capable employees, and lowering total costs.
- A sensitivity analysis with respect to the sampling variance, σ^2, indicates that lower values of σ

result in a smaller fraction of employees who are terminated. Thus, reductions in within-period variability improve the selectivity and effectiveness of screening procedures, allowing the employer to reduce termination rates obtained using the optimal policy.

• When training costs are absent, the screening process is very selective, terminating more than half of employees on day 1 and more than 85% of employees overall. When training costs are present, however, termination rates decrease as training costs increase.

For additional details concerning these results, contact the authors.

7. Conclusions

This paper studies how statistical and on-the-job learning together determine the nature of optimal hiring and retention decisions. Statistical learning arises when workers are heterogeneous and the employer does not know their true quality. On-the-job learning occurs as experience affects workers' performance.

The literature related to this problem comes from various areas, such as labor economics, statistical decision theory, learning-curve theory, and service operations, among others. Our analysis integrates aspects from all of these streams to incorporate training, switching, and quitting dynamics, and it applies results from infinite-armed Bayesian bandit problems to characterize the optimal hiring and retention policies.

Our numerical results show that active screening of employees can significantly improve expected costs and long-run average employee performance. Because most termination takes place early in employees' tenures, relatively simple finite-horizon and one-shot policies also have the potential to perform well. Our sensitivity analysis shows that, as is common in bandit problems, the ability to terminate employees should motivate managers to consider a broader spectrum of potential hires. Moreover, both reductions in within-task variability and improvements in employee learning provide the additional benefit of lowering termination rates.

Acknowledgments

The authors are grateful to an associate editor and three anonymous referees whose comments helped to significantly improve the paper. The authors also gratefully acknowledge financial support from the Wharton–INSEAD Alliance's Center for Global Research and Education. This material is based on work supported by the National Science Foundation [Grant CMMI-0800645]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Appendix. Mathematical Results

Proofs of mathematical claims are presented in the order of their appearance in the main paper. When other technical results are needed, they are stated with a full proof or suitable reference in the location that they are needed.

To simplify the exposition, we introduce the following shorthand. For any given initial state, $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, let $M_i \equiv M(\nu_{i,0}, n_{i,0}, u_{i,0})$ denote the initial value of worker i's index, $\tilde{\Lambda}_i(m) \equiv \tilde{\Lambda}(\nu_{i,0}, n_{i,0}, u_{i,0}, m)$ be the stopping time (15), $HV_i(m) \equiv HV(\nu_{i,0}, n_{i,0}, u_{i,0}, m)$ be i's expected continuation cost (16), and $C_{i,t} \equiv c_s u_{i,t} + c_h \mathbb{1}(n_{i,t} = 0) + c(Z(\nu_{i,t}, n_{i,t})) + (c_q - c_s)\mathbb{1}(\nu_{\pi(t-1),t} = \mathbb{I}_K \cap t > 0)$ be worker i's one-period cost for being employed at time t.

Proof of Lemma 1

For (i), let π be a hiring and retention policy for Problem 2 that employs unproductive workers. Then, let $T = \inf\{t\colon n_{\pi(t),t} \ge \Lambda_{\pi(t)}\}$ be the first time that such a worker is employed. The one-period cost at time T for employing the unproductive worker $\pi(T)$ is $\gamma^T K$. Construct a new policy π^T such that $\pi^T(t) = \pi(t)$ for $t < T$, and $\pi^T(t) = \pi(t+1)$ for $t \ge T$. For any initial state $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, we have that

$$C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = -c_s + \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t C_{\pi(t),t} + \gamma^T K + \sum_{t=T+1}^{\infty} \gamma^t C_{\pi(t),t}\right],$$

$$C^0_{\pi^T}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = -c_s + \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t C_{\pi(t),t} + \sum_{t=T+1}^{\infty} \gamma^{t-1} C_{\pi(t),t}\right],$$

and

$$\begin{aligned} &C^0_{\pi^T}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &\quad = \mathbb{E}\left[(1-\gamma)\left(\sum_{t=T+1}^{\infty} \gamma^{t-1} C_{\pi(t),t}\right) - \gamma^T K\right] \\ &\quad < \mathbb{E}\left[(1-\gamma)\left(\sum_{t=T+1}^{\infty} \gamma^{t-1} K\right) - \gamma^T K\right] \\ &\quad = \mathbb{E}\left[(1-\gamma)\frac{\gamma^T K}{1-\gamma} - \gamma^T K\right] = 0. \end{aligned}$$

Thus, the infinite-horizon total expected discounted cost of π^T is strictly smaller than that of π, and π cannot be optimal.

For (ii) we begin with the *if* part.

If: Let π^* be optimal for Problem 2. Then by part (i) of the lemma, policy π^* employs no unproductive worker; therefore, π^* is feasible for Problem 1. For any initial state $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ and for any policy π feasible for Problem 1, we let $C^0_{\pi,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ and $C^0_{\pi,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, respectively, be the infinite-horizon total expected discounted cost of policy π in Problems 1 and 2, and we observe that $C^0_{\pi,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$. Because π^* is optimal for Problem 2 and feasible for Problem 1, we obtain that $C^0_{\pi^*,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi^*,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ for all π feasible for Problem 1. Hence, π^* is also optimal for Problem 1.

Only if: Let π^* be optimal for Problem 1. Then, any policy, π, that is feasible for Problem 1 is feasible for Problem 2, and $C^0_{\pi,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$. By part (i) of the lemma, we know that any policy π that is feasible for Problem 2 but not for Problem 1 cannot be optimal. Then, by optimality and feasibility, we obtain that $C^0_{\pi^*,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi^*,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi,1}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi,2}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, and policy π^* is also optimal for Problem 2. □

Proof of Lemma 2

Suppose that $\pi \in \Pi$ is a policy for Problem 1 and that $\mathbb{E}[\Lambda_i] < \infty$ for all $i \in \mathcal{S}_0$. No policy for Problem 1 can use an employee after he has quit. Thus, the random variable $\Lambda_i(\pi)$ in (3) satisfies $0 \le \Lambda_i(\pi) \le \Lambda_i$ on every sample path, for all $i \in \mathcal{S}_0$. Suppose, by contradiction, that policy $\pi \in \Pi$ only employs $\kappa < \infty$ workers with some positive probability $\epsilon > 0$. Because $\pi \in \Pi$, we have

$$\mathbb{P}\left(\sum_{i=1}^{\kappa} \Lambda_i(\pi) \ge \zeta\right) \ge \epsilon \tag{21}$$

for all $\zeta \in \mathbb{R}$. Given that $0 \le \Lambda_i(\pi) \le \Lambda_i$ and that the Λ_i's are iid, Markov's inequality implies that $\mathbb{P}(\sum_{i=1}^{\kappa} \Lambda_i(\pi) \ge \zeta) \le \mathbb{P}(\sum_{i=1}^{\kappa} \Lambda_i \ge \zeta) \le \kappa\mathbb{E}[\Lambda_1]/\zeta$. Picking any $\zeta > \kappa\mathbb{E}[\Lambda_1]/\epsilon$ would contradict (21) from which we conclude that $\pi \notin \Pi$. Hence, each policy for Problem 1 employs an infinite number of workers with probability 1. □

Properties of the Value Function and of the Gittins Index

Lemma 3. *For each ν, n, and u, $V(\nu, n, u, m)$ is concave, nondecreasing, and Lipschitz continuous in m, with Lipschitz constant equal to* 1.

Proof. We proceed by means of the value iteration algorithm (see, e.g., Bertsekas and Shreve 1978, §9.5, Definition 9.10, and Proposition 9.14). Let $v^0(\nu, n, u, m) = 0$ for all $m \in \mathbb{R}$, and note that v^0 is trivially nondecreasing, concave, and Lipschitz-1 continuous in m for each ν, n, and u. Assume that $v^{k-1}(\nu, n, u, m)$ is nondecreasing, concave, and Lipschitz-1 continuous in m for each ν, n, and u. Let

$$\begin{aligned} v^k&(\nu, n, u, m)\\ &= \min\{m, c_s u + c_h \mathbb{1}(n=0) + \mathbb{E}[c(Z(\nu, n))]\\ &\qquad + \gamma(1-q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), n+1, 0, m)]\\ &\qquad + \gamma q_n[c_q - c_s + v^{k-1}(\mathbb{1}_K, n+1, 0, m)]\}, \end{aligned}$$

and note that $c_s u + c_h \mathbb{1}(n=0) + \mathbb{E}[c(Z(\nu, n))]$ is constant with respect to m, and $\gamma(1-q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), n+1, 0, m)]$ is nondecreasing, concave, and Lipschitz-$\gamma(1-q_n)$ continuous in m by the induction assumption and the fact that these properties are preserved when taking expectations. The induction assumption also yields that $\gamma q_n[c_q - c_s + v^{k-1}(\mathbb{1}_K, n+1, 0, m)]$ is nondecreasing, concave, and Lipschitz-γq_n continuous in m. Monotonicity and concavity are preserved under minimization, so we have that $v^k(\nu, n, u, m)$ is nondecreasing and concave in m.

To obtain that $v^k(\nu, n, u, m)$ is also Lipschitz-1 continuous in m, the argument is similar, but a little more care is required. Given two Lipschitz functions h, h' with Lipschitz constants c_1, c_2, respectively, $\min\{h, h'\}$ is Lipschitz with constant $c_3 = \max\{c_1, c_2\}$. In our context, the left minimand is Lipschitz-1 continuous, and the right minimand is Lipschitz-γ continuous, with $\gamma < 1$, so that $v^k(\nu, n, u, m)$ is also Lipschitz-1 continuous in m. To conclude our argument, we let $k \to \infty$ so $v^k(\nu, n, u, m) \to V(\nu, n, u, m)$. □

Lemma 4. (i) *$HV(\nu, n, u, m) < m$ if and only if $M(\nu, n, u) < m$.*

(ii) *$HV(\nu, n, u, m) > m$ if and only if $m < M(\nu, n, u)$.*

(iii) *$HV(\nu, n, u, m) = m$ if and only if $m = M(\nu, n, u)$.*

Proof. We prove each of the three statements in turn.

(i) If $M(\nu, n, u) < m$, then $V(\nu, n, u, m) < m$. In turn, $V(\nu, n, u, m) < m$ implies it is optimal not to retire so $HV(\nu, n, u, m) = V(\nu, n, u, m) < m$. If $HV(\nu, n, u, m) < m$, we have that $HV(\nu, n, u, m) = V(\nu, n, u, m) < m$. Then the fact that $M(\nu, n, u) < m$ follows by the definition of the Gittins index (17), $M(\nu, n, u)$, and the fact that the Bellman Equation (12), $V(\nu, n, u, m)$, is concave and nondecreasing in m with $V(\nu, n, u, m) \le m$ for all m.

(ii) It follows directly from the proof of (i) by reversing the inequalities.

(iii) It follows combining claims (i) and (ii). □

Lemma 5. *For each ν and n, $M(\nu, n, 0) \le M(\nu, n, 1)$.*

Proof. Because $c_s \ge 0$, it is immediate to see that $V(n, \nu, 0, m) \le V(n, \nu, 1, m)$ for each m. Then, given the monotonicity property of the value function $V(\nu, n, u, m)$ in m for each given n, ν, and u (Lemma 3), we have that $M(\nu, n, 0) = \sup\{m\colon V(n, \nu, 0, m) = m\} \le \sup\{m\colon V(n, \nu, 1, m) = m\} = M(n, \nu, 1)$. □

Proof of Proposition 1, Part (i)

To prove Proposition 1, part (i), we first prove, in Lemma 6, the optimality of an index policy in the case that $c_s = 0$. The lemma is an analogue of Theorem 1 in Bergemann and Välimäki (2001) and Theorem 4.2 in Sundaram (2005). Its proof follows along the lines argued in Sundaram (2005).

Lemma 6. *If conditions of Proposition* 1 *hold and $c_s = 0$, then a policy π^* is optimal if and only if*

$$\pi^*(t) \in \left\{ i \in \mathcal{S}_0\colon M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) = \inf_{j \in \mathcal{S}_0} M_j(\nu_{j,t}, n_{j,t}, u_{j,t}) \right\},$$
$$\textit{a.s. for all } t = 0, 1, 2, \ldots.$$

Proof. Given the initial state $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ such that $\nu_{i,0} \equiv \hat{\nu}$, $n_{i,0} \equiv 0$, and $u_{i,0} \equiv 1$ for all $i \in \mathcal{S}_0$, we have that all workers have index $M(\hat{\nu}, 0, 1) \equiv \hat{m}$. Thus, at any time t there are at most t workers that have been employed, so that there are at most t indices with value different than $\hat{m}$. Hence, for each $t = 0, 1, 2, \ldots$, the infimum in Lemma 6 is attained, and the index policy described in Lemma 6 is well defined.

Because $c_s = 0$, the optimality equation (12) and the expected discounted cost of continuing (13) are constant with respect to u, i.e., $V(\nu, n, 0, m) = V(\nu, n, 1, m)$ and $HV(\nu, n, 0, m) = HV(\nu, n, 1, m)$ for all ν, n, m. We also have $M(\nu, n, 0) = M(\nu, n, 1)$, and the value of the Gittins index of a given worker is independent from that of other workers.

To prove Lemma 6, we now introduce some additional notation. We let $\pi(j)$ be the hiring and retention policy that begins by employing worker j and continues according to the index rule. We also let $\pi(i, j)$ be the policy that first employs worker i (with ability distribution $\nu_{i,0}$, experience $n_{i,0}$, and switching indicator $u_{i,0}$) as long as his Gittins index does not exceed its original value, $M(\nu_{i,0}, n_{i,0}, u_{i,0})$. Policy $\pi(i, j)$ then employs worker j for at least one period, until j's index exceeds the original value of worker i's index, $M(\nu_{i,0}, n_{i,0}, u_{i,0})$. After employing worker i and j as described, policy $\pi(i, j)$ continues according to the index rule.

Lemmas 7–9 study the cost of the employment policies $\pi(i, j)$, $\pi(j, i)$, $\pi(i)$, and $\pi(j)$. The lemmas hold for any initial state, $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, such that there are infinitely many workers, i, with $\nu_{i,0} = \hat{\nu}$, $n_{i,0} = 0$, and $u_{i,0} = 1$.

Lemma 7. *If $M_i = M_j$, then $C^0_{\pi(i,j)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = C^0_{\pi(j,i)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$.*

Proof. By construction, the infinite-horizon expected discounted cost of policy $\pi(i, j)$ is

$$\begin{aligned} C^0_{\pi(i,j)}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &= \mathbb{E}\Bigg[\sum_{t=0}^{\tilde{\Lambda}_i(M_i)-1} \gamma^t C_{i,t} + \sum_{t=\tilde{\Lambda}_i(M_i)}^{\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)-1} \gamma^t C_{j,t} \\ &\qquad + \sum_{t=\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)}^{\infty} \gamma^t C_{\pi(t),t}\Bigg] \\ &= HV_i(M_i) + \mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_i)}]\{-M_i + HV_j(M_i) - \mathbb{E}[\gamma^{\tilde{\Lambda}_j(M_i)}]M_i\} \\ &\quad + C^{\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)}_{\pi(i,j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}'), \end{aligned} \tag{22}$$

where $C^{\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)}_{\pi(i,j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ is the expected discounted (to $t = 0$) continuation cost of policy $\pi(i, j)$ after having employed worker i for $\tilde{\Lambda}_i(M_i)$ periods and worker j for $\tilde{\Lambda}_j(M_i)$ periods. Because only workers i and j have been employed, the new state, $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$, differs from $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ only in its ith and jth coordinates. Similarly,

$$\begin{aligned} C^0_{\pi(j,i)}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &= HV_j(M_j) + \mathbb{E}[\gamma^{\tilde{\Lambda}_j(M_j)}]\{-M_j + HV_i(M_j) - \mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_j)}]M_j\} \\ &\quad + C^{\tilde{\Lambda}_j(M_j)+\tilde{\Lambda}_i(M_j)}_{\pi(j,i)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}'). \end{aligned}$$

Because $M_i = M_j$, we have that, at time $\tilde{\Lambda}_i(M_i) + \tilde{\Lambda}_j(M_i)$, the continuation costs $C^{\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)}_{\pi(i,j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ and $C^{\tilde{\Lambda}_j(M_j)+\tilde{\Lambda}_i(M_j)}_{\pi(j,i)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ are equal. Moreover, we can use Lemma 4 to obtain that $HV_j(M_i) = M_i = HV_i(M_i)$ and $HV_i(M_j) = M_j = HV_j(M_j)$ so that

$$\begin{aligned} C^0_{\pi(i,j)}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi(j,i)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &= M_i - \mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_i)}]\mathbb{E}[\gamma^{\tilde{\Lambda}_j(M_i)}]M_i - M_i + \mathbb{E}[\gamma^{\tilde{\Lambda}_j(M_i)}]\mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_i)}]M_i \\ &= 0. \quad \square \end{aligned}$$

Lemma 8. *If $M_i = \inf_k M_k$ and $M_i < M_j$, then $C^0_{\pi(i,j)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) < C^0_{\pi(j)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$.*

Proof. Policy $\pi(j)$ employs worker j for the first period and then continues according to the index rule. After his first performance, worker j is retained as long as he is index minimal. When worker j is terminated, Lemma 7 shows us that we can choose policy $\pi(j)$ to employ worker i and continue with the index rule. Thus,

$$\begin{aligned} C^0_{\pi(j)}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &= HV_j(M_i) + \mathbb{E}[\gamma^{\tilde{\Lambda}_j(M_i)}]\{-M_i + HV_i(M_i) - \mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_i)}]M_i\} \\ &\quad + C^{\tilde{\Lambda}_j(M_i)+\tilde{\Lambda}_i(M_i)}_{\pi(j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}'), \end{aligned}$$

where $C^{\tilde{\Lambda}_j(M_i)+\tilde{\Lambda}_i(M_i)}_{\pi(j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ is the expected discounted continuation cost of policy $\pi(j)$ after having employed worker j for $\tilde{\Lambda}_j(M_i)$ periods and worker i for $\tilde{\Lambda}_i(M_i)$ periods. The new state, $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$, differs from $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ only in its jth and ith coordinates.

We now recall the representation (22) for the expected cost of policy $\pi(i, j)$, and we observe that the expected continuation costs $C^{\tilde{\Lambda}_i(M_i)+\tilde{\Lambda}_j(M_i)}_{\pi(i,j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ and $C^{\tilde{\Lambda}_j(M_i)+\tilde{\Lambda}_i(M_i)}_{\pi(j)}(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$ are equal. From Lemma 4, we know that $HV_i(M_i) = M_i$. Because $M_i < M_j$, we also have $M_i < HV_j(M_i)$. Then,

$$\begin{aligned} C^0_{\pi(i,j)}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi(j)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \\ &= [M_i - HV_j(M_i)](1 - \mathbb{E}[\gamma^{\tilde{\Lambda}_i(M_i)}]) < 0. \quad \square \end{aligned}$$

Lemma 9. *If $M_i = \inf_k M_k$ and $M_i < M_j$, then $C^0_{\pi(i)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) < C^0_{\pi(j)}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$.*

Proof. Because $M_i = \inf_k M_k$ and $M_i < M_j$, Lemma 8 shows us that policy $\pi(i, j)$ strictly improves policy $\pi(j)$. We now argue that $\pi(i, j)$ can be improved by employing a Gittins-index-minimal worker any time that is not prescribed. The first worker that is employed by policy $\pi(i, j)$, i, is Gittins-index minimal. At his termination, the state of the system changes from the initial $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ to $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$, which differs only in the ith coordinate. After the employment of worker i, policy $\pi(i, j)$ prescribes the employment of worker j. Its continuation value then equals that of policy $\pi(j)$ when starting in state $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$. Lemma 8 then shows us that if j is not Gittins-index minimal at $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$, then it is strictly better to use the policy $\pi(l, j)$, where worker l is such that $M_l = \inf_k M_k$ and $M_l < M_j$. Iterating on this reasoning, we obtain that policy $\pi(i)$, the index policy, is strictly better than any index policy that begins with a worker that is not index minimal. $\square$

We are now ready to complete the proof of Lemma 6.

If: Let π be any employment policy and consider the policy π^T such that $\pi^T(t) = \pi(t)$ for all $0 \le t < T$ and $\pi^T(t) = \pi^*(t)$ for $T \le t$, where π^* denotes the index rule. At any time T, the system is in state $(\boldsymbol{\nu}', \mathbf{n}', \mathbf{u}')$, which is different from the initial $(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ in at most T coordinates. Thus, there are infinitely many workers whose state has never changed, and whose index equals $\hat{m}$, so that policy π^T is well defined. Because the problem is discounted $(\gamma < 1)$ and the one-period costs are uniformly bounded, we can pick any $\epsilon > 0$ and choose T so that $C^0_{\pi^T}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) < \epsilon$. Then, according to Lemma 9, we might improve policy π^T by employing a Gittins-index-minimal worker at time $T-1$. Thus, $C^0_{\pi^{T-1}}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi^T}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ and also $C^0_{\pi^{T-1}}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) < \epsilon$. Iterating back to $T = 1$, we have that $C^0_{\pi^0}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) - C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) < \epsilon$, where π^0 is the index policy π^*. Because ϵ is arbitrary, we then have that $C^0_{\pi^0}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$. Because the choice of policy π was also arbitrary, we can choose π to be any optimal policy so that $C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi^0}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) \le C^0_{\pi}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$. Thus, the index policy π^0 is optimal too.

Only if: Let π be an optimal policy and assume that π is *not* an index policy. Let T be the first time at which π

does not employ a Gittins-index-minimal worker, and construct the policy $\hat{\pi}$ such that $\hat{\pi}(t) = \pi(t)$ for all $0 \le t \le T$ and $\hat{\pi}(t) = \pi^*(t)$ for all $T < t$, where, as usual, π^* denotes the index policy. Because both π and π^* are optimal, policy $\hat{\pi}$ is optimal too. However, by Lemma 9 we can strictly improve on policy $\hat{\pi}$ by selecting an index-minimal worker at time T, and by doing so we obtain that policies $\hat{\pi}$ and π cannot be optimal, a contradiction. □

Having proved the optimality of a Gittins-index policy when $c_s = 0$, we now prove Proposition 1, part (i), which allows $c_s > 0$. The proof's argument follows the sketch of Theorem 2 provided in Bergemann and Välimäki (2001).

Proof of Proposition 1, Part (i). Let $C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_h, c_s)$ be the cost function (8) that makes explicit the dependence on the training cost, c_h, and on the switching cost, c_s. We know from Lemma 6 that $C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_h, 0)$ is minimized if and only if π is an index policy. Similarly, the same happens for $C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_s + c_h, 0)$ because we are just imposing a different training cost, $c_s + c_h$. For all policy $\pi \in \Pi$, we then have that

$$C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_s + c_h, 0) \le C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_s + c_h, 0) \le c_s + C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_h, c_s). \quad (23)$$

The first inequality holds by the optimality of policy π^*. The second inequality holds because the switching cost, c_s, is incurred every time the workers employed in two subsequent periods differ (not only at the first employment of a new worker). The second inequality is met with equality for all policies π that never recall previously employed workers.

If: We now show that if π is the index policy in Proposition 1, part (i), then $c_s + C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ achieves the lower bound (23). At time $t = 0$, all workers have the same index, $\hat{m}$, and the employer chooses a worker, i, at random from the pool. Worker i is then employed for $\tilde{\Lambda}_i(\hat{m})$ periods, and his index $M_i(\nu_{i,\tilde{\Lambda}_i(\hat{m})}, n_{i,\tilde{\Lambda}_i(\hat{m})}, 0) > \hat{m}$. Because worker i is not index minimal at time $\tilde{\Lambda}_i(\hat{m})$, another worker, j, is employed. This causes a transition of the state of worker i, from $(\nu_{i,\tilde{\Lambda}_i(\hat{m})}, n_{i,\tilde{\Lambda}_i(\hat{m})}, 0)$ to $(\nu_{i,\tilde{\Lambda}_i(\hat{m})+1}, n_{i,\tilde{\Lambda}_i(\hat{m})+1}, 1)$, with $\nu_{i,\tilde{\Lambda}_i(\hat{m})} = \nu_{i,\tilde{\Lambda}_i(\hat{m})+1}$, and $n_{i,\tilde{\Lambda}_i(\hat{m})} = n_{i,\tilde{\Lambda}_i(\hat{m})+1}$. By Lemma 5 we know that $M(n, \nu, 0) \le M(n, \nu, 1)$ for each ν, n. Because worker i in state $(\nu_{i,\tilde{\Lambda}_i(\hat{m})}, n_{i,\tilde{\Lambda}_i(\hat{m})}, 0)$ has index exceeding $\hat{m}$, the same happens to worker i when in state $(\nu_{i,\tilde{\Lambda}_i(\hat{m})+1}, n_{i,\tilde{\Lambda}_i(\hat{m})+1}, 1)$.

Repeating this argument for all employed workers, we see that the transition of u from 0 to 1 only increases the indices of workers whose indices are greater than $\hat{m}$ and, in turn, does not change the dynamics of the index policy, which then agrees with the index policy, π^*, used to achieve $C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_s + c_h, 0)$.

Only if: Assume that π is an optimal policy for $C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_h, c_s)$. From the *if* part of the proof, we know that an optimal π satisfies

$$C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_s + c_h, 0) = c_s + C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}, c_h, c_s);$$

i.e., it achieves the lower bound (23). Then π is also an optimal policy for $C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$, and Lemma 6 shows us that π must be an index policy. □

Proof of Proposition 1, Part (ii)
This result is an analogue of Corollary 1 in Bergemann and Välimäki (2001).

At $t = 0$, no worker has ever been employed and all the workers have Gittins index $\hat{m}$. Then, the sampling process starts with a random selection of worker i from the stationary pool of candidates. Worker i is employed at all times, t, such that $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) = \inf_j\{M_j(\nu_{j,t}, n_{j,t}, u_{j,t})\} \le \hat{m}$. As soon as i is discarded, $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) > \hat{m}$ and the sampling process starts again. □

Proof of Proposition 1, Part (iii)
The result follows immediately from Lemma 1 and Proposition 1, part (i). □

Proof of Proposition 1, Part (iv)
This result is an analogue of (2) in Bergemann and Välimäki (2001).

Consider the retirement-option problem described in §3. By Lemma 4(iii), we obtain $\hat{m} = HV(\hat{\nu}, 0, 1, \hat{m})$, and we note that $HV(\hat{\nu}, 0, 1, \hat{m})$ is the total expected discounted cost of employing a productive worker, i, with ability distribution $\nu_{i,0} = \hat{\nu}$, experience $n_{i,0} = 0$, and switching indicator $u_{i,0} = 1$ for at least one period followed by an optimal termination decision that depends on the retirement payment $\hat{m}$. Recall now the definition of the optimal stopping time $\tilde{\Lambda}(\nu, n, u, m)$ in (15) and the stopping-time representation for $HV(\nu, n, u, m)$ in (16). Thus,

$$HV(\hat{\nu}, 0, 1, \hat{m}) = \mathbb{E}\bigg[c_s + c_h + \sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1} \gamma^r c(Z(\nu_r, r)) + \gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}[(c_q - c_s)\mathbb{1}_{Q_0} + \hat{m}]\bigg].$$

Because $\hat{m} = HV(\hat{\nu}, 0, 1, \hat{m})$, we obtain

$$(1 - \mathbb{E}[\gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}])\hat{m} = \mathbb{E}\bigg[c_s + c_h + \sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1} \gamma^r c(Z(\nu_r, r)) + \gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}(c_q - c_s)\mathbb{1}_{Q_0}\bigg]. \quad (24)$$

At time $t = 0$ all workers $i \in \mathcal{S}_0$ have ability distribution $\nu_{i,0} = \hat{\nu}$, experience $n_{i,0} = 0$, and switching indicator $u_{i,0} = 1$. Lemma 6 shows us that worker i can be optimally retained at time t if and only if his Gittins index is minimal, i.e., $M_i(\nu_{i,t}, n_{i,t}, u_{i,t}) \le \hat{m}$. Worker i stops being employed at time $\tilde{\Lambda}_i(\hat{\nu}, 0, 1, \hat{m})$, either because he is terminated or he quits. Because all workers $i \in \mathcal{S}_0$ are identical, the sequence $\{\tilde{\Lambda}_i \equiv \tilde{\Lambda}_i(\hat{\nu},0,1,\hat{m}),\ i = 1, 2, 3, \ldots\}$ is iid. Set $\tilde{\Lambda}_0 \equiv 0$, recall that Λ_i is the time at which worker i becomes unproductive, and let $Q_{i,0} = \{\omega\colon \tilde{\Lambda}_i(\hat{\nu}, 0, 1, \hat{m}) = \Lambda_i\}$ be the set of sample paths for which worker i quits before he is terminated. Then,

$$\inf_{\pi\in\Pi} C^0_\pi(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = -c_s + \mathbb{E}\bigg[\sum_{k=1}^{\infty} \gamma^{\sum_{i=0}^{k-1}\tilde{\Lambda}_i} \sum_{r=0}^{\tilde{\Lambda}_k-1} \gamma^r\{(c_s + c_h)\mathbb{1}(r = 0) + c(Z(\nu_r, r))\} + \gamma^{\tilde{\Lambda}_k}(c_q - c_s)\mathbb{1}_{Q_{k,0}}\bigg]$$

$$= -c_s + \sum_{k=1}^{\infty} \mathbb{E}[\gamma^{\sum_{i=0}^{k-1}\tilde{\Lambda}_i}]\mathbb{E}\bigg[\sum_{r=0}^{\tilde{\Lambda}_k-1}\gamma^r\{(c_s+c_h)\mathbb{1}(r=0)+c(Z(\nu_r,r))\} + \gamma^{\tilde{\Lambda}_k}(c_q-c_s)\mathbb{1}_{Q_{k,0}}\bigg]$$

$$= -c_s + \hat{m}(1-\mathbb{E}[\gamma^{\tilde{\Lambda}_1}])\sum_{k=1}^{\infty}\mathbb{E}[\gamma^{\sum_{j=0}^{k-1}\tilde{\Lambda}_j}] = -c_s + \hat{m}, \tag{25}$$

where (25) follows from (24), and $\sum_{k=1}^{\infty}\mathbb{E}[\gamma^{\sum_{j=0}^{k-1}\tilde{\Lambda}_j}] = 1 + \mathbb{E}[\gamma^{\tilde{\Lambda}_1}] + \mathbb{E}[\gamma^{\tilde{\Lambda}_1}]^2 + \cdots = (1-\mathbb{E}[\gamma^{\tilde{\Lambda}_1}])^{-1}$. □

Proof of Proposition 2

We first prove the following lemma that uses the notion of a likelihood-ratio (lr) order (Shaked and Shanthikumar 2007, §1.C). Suppose that X is a random variable with probability density function (pdf) f_X and that Y is a random variable with pdf f_Y. We write $X \le_{\mathrm{lr}} Y$ (X is stochastically smaller than Y in the likelihood ratio sense) if $f_Y(z)/f_X(z)$ increases in z over the union of the supports of X and Y.

Lemma 10. *Let g: $\mathbb{R}^3 \to \mathbb{R}$ be such that for $A \sim \nu$, $\beta(\nu, z)([-\infty, a]) = \mathbb{P}(A \le a \mid Z = z)$ is nondecreasing in z for any given ν. If, for any $a \le a'$, $\xi_n(z \mid a')/\xi_n(z \mid a)$ is nondecreasing in z, then $V(\nu, n, u, m) \le V(\nu', n, u, m)$ for any $\nu \le_{\mathrm{lr}} \nu'$, and for each given n, u, m.*

The monotonicity of the Bayes operator ensures that the Bayesian update implies that larger observations lead to stochastically larger posterior distributions in some sense. Note also that for several well-known families of distributions, the likelihood-ratio comparison can be simply checked by comparing distribution parameters. Müller and Stoyan (2002, Table 1.1) propose such comparison criteria for several continuous and discrete distributions.

Proof of Lemma 10. To show monotonicity of the value function (12) with respect to the likelihood-ratio order, we proceed by means of the value iteration algorithm (see, e.g., Bertsekas and Shreve 1978, §9.5, Definition 9.10, and Proposition 9.14). We fix n, u, and m, and we let $v^0(\nu, n, u, m) = 0$ for all distributions ν. Trivially, we have that v^0 is lr-nondecreasing in ν. We then assume that $v^{k-1}(\nu, n, u, m) \le v^{k-1}(\nu', n, u, m)$ for $\nu \le_{\mathrm{lr}} \nu'$, and we write

$$\begin{aligned} &v^k(\nu, n, u, m) \\ &\quad = \min\big\{m,\ c_s u + c_h\mathbb{1}(n=0) + \mathbb{E}[c(Z(\nu, n))] \\ &\qquad + \gamma(1-q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), n+1, 0, m)] \\ &\qquad + \gamma q_n[c_q - c_s + v^{k-1}(\mathbb{1}_K, n+1, 0, m)]\big\}. \end{aligned} \tag{26}$$

In Equation (26), we first notice that the quantity $c_s u + c_h\mathbb{1}(n=0) + \gamma q_n[c_q - c_s + v^{k-1}(\mathbb{1}_K, n+1, 0, m)]$ is independent, and hence constant, with respect to ν and ν'.

Because $\xi_n(z \mid a')/\xi_n(z \mid a)$ is nondecreasing in z for any $a \le a'$, the definition of the likelihood-ratio order yields that $Z(a, n) \le_{\mathrm{lr}} Z(a', n)$. From Shaked and Shanthikumar (2007, Theorem 1.C.17) and the fact that $\nu \le_{\mathrm{lr}} \nu'$, we obtain that $Z(\nu, n) \le_{\mathrm{lr}} Z(\nu', n)$. We now also have that $\mathbb{E}[Z(\nu, n)] \le \mathbb{E}[Z(\nu', n)]$ because the likelihood-ratio order ($\le_{\mathrm{lr}}$) implies the usual stochastic order ($\le_{\mathrm{st}}$) (Shaked and Shanthikumar 2007, Theorem 1.C.1).

Finally, by noting that $Z(\nu, n) \le_{\mathrm{lr}} Z(\nu', n)$, we obtain that

$$\beta(\nu, Z(\nu, n)) \le_{\mathrm{lr}} \beta(\nu, Z(\nu', n)) \le_{\mathrm{lr}} \beta(\nu', Z(\nu', n)).$$

The first ordering holds because $\beta(\nu, z)$ is nondecreasing in z (Shaked and Shanthikumar 2007, Theorem 1.C.8). The second ordering holds because $\nu \le_{\mathrm{lr}} \nu'$ (Shaked and Shanthikumar 2007, Example 1.C.58). Then, the induction assumption and the monotonicity property of the expected value yield that $\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), n+1, 0, m)] \le \mathbb{E}[v^{k-1}(\beta(\nu', Z(\nu', n)), n+1, 0, m)]$. Thus, the right minimand in (26) is lr-nondecreasing in ν. The first minimand, m, is constant with respect to ν, so that $v^k(\nu, n, u, m) \le v^k(\nu', n, u, m)$, provided that $\nu \le_{\mathrm{lr}} \nu'$. Repeated application of the value iteration algorithm then yields $V(\nu, n, u, m) \le V(\nu', n, u, m)$, for any $\nu \le_{\mathrm{lr}} \nu'$, as desired. □

Proof of Proposition 2. The posterior distribution of A has distribution $N(w_p, \sigma^2/p)$. The normal distribution has the monotone likelihood-ratio property required by Lemma 10 (see, e.g., Müller and Stoyan 2002, Table 1.1). An application of that lemma proves the desired monotonicity for V.

For the Gittins index we have the following. Given $\nu \sim N(w_p, \sigma^2/p)$ and $\nu' \sim N(w'_p, \sigma^2/p)$ with $w_p \le w'_p$, we have that $\nu \le_{\mathrm{lr}} \nu'$, so $V(\nu, n, u, m) \le V(\nu', n, u, m)$ for any n, u, and m. Then,

$$\begin{aligned} M(\nu, n, u) &= \sup\{m\colon V(\nu, n, u, m) = m\} \\ &\le \sup\{m\colon V(\nu', n, u, m) = m\} = M(\nu', n, u). \quad \square \end{aligned}$$

Proof of Theorem 1

For each given ν, n, u, and m, recall the definition of the stopping time $\tilde{\Lambda}(\nu, n, u, m)$ in (15). Also, recall that Λ_i is the time at which i becomes unproductive, and let $Q_{i,0} = \{\omega\colon \tilde{\Lambda}_i(\hat{\nu}, 0, 1, \hat{m}) = \Lambda_i\}$ be the set of sample paths for which worker i with state $(\hat{\nu}, 0, 1)$ quits before he is terminated. At time $t = 0$ all workers $i \in \mathcal{S}_0$ have $\nu_{i,0} = \hat{\nu}$, $n_{i,0} = 0$, and $u_{i,0} = 1$. From the proof of Proposition 1, part (iv), we know that the sequence $\{\tilde{\Lambda}_i(\hat{\nu}, 0, 1, \hat{m}),\ i = 1, 2, 3, \ldots\}$ is iid. The $\{\Lambda_i,\ i = 1, 2, 3, \ldots\}$ are also iid, so that the $Q_{i,0}$'s are iid too.

Because $q_{i,n} = q$ for all $i \in \mathcal{S}_0$ and all n, the proof of this result hinges on showing that

$$\mathbb{E}\bigg[\sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1}\gamma^{r+1}q\bigg] = \mathbb{E}[\gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}\mathbb{1}_{Q_0}], \tag{27}$$

which would imply that

$$\begin{aligned} &\mathbb{E}\bigg[\sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1}\gamma^r c(Z(\nu_r, r)) + \gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}(c_q - c_s)\mathbb{1}_{Q_0}\bigg] \\ &\quad = \mathbb{E}\bigg[\sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1}\gamma^r\{c(Z(\nu_r, r)) + \gamma q(c_q - c_s)\}\bigg]. \end{aligned} \tag{28}$$

This would give an alternative representation for $C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$. Under the optimal employment policy,

$$C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u}) = -c_s + \mathbb{E}\bigg[\sum_{k=1}^{\infty}\gamma^{\sum_{i=0}^{k-1}\tilde{\Lambda}_i}\sum_{r=0}^{\tilde{\Lambda}_k-1}\gamma^r\big\{(c_s+c_h)\mathbb{1}(r=0) + c(Z(\nu_r, r))\big\} + \gamma^{\tilde{\Lambda}_k}(c_q - c_s)\mathbb{1}_{Q_{i,0}}\bigg],$$

where the $\tilde{\Lambda}_k \equiv \tilde{\Lambda}_k(\hat{\nu}, 0, 1, \hat{m})$ for all k, $\tilde{\Lambda}_0 \equiv 0$. Then, (28) allows us to write $C^0_{\pi^*}(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})$ as

$$\begin{aligned}C^0_{\pi^*}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})\\ &= -c_s + \mathbb{E}\bigg[\sum_{k=1}^{\infty} \gamma^{\sum_{i=0}^{k-1}\tilde{\Lambda}_i} \sum_{r=0}^{\tilde{\Lambda}_k - 1} \gamma^r \big\{(c_s + c_h)\mathbb{1}(r=0)\\ &\qquad + c(Z(\nu_r, r)) + \gamma q(c_q - c_s)\big\}\bigg]\\ &= -c_s + \inf_{\pi\in\Pi}\bigg\{\mathbb{E}\bigg[\sum_{t=0}^{\infty}\gamma^t\big\{(c_s+c_h)\mathbb{1}(n_{\pi(t),t}=0)\\ &\qquad + c(Z(\nu_{\pi(t),t}, n_{\pi(t),t})) + \gamma q(c_q - c_s)\big\}\bigg]\bigg\}. \quad (29)\end{aligned}$$

The quantity $\gamma q(c_q - c_s)$ is a shifting constant that does not affect the minimization problem, so we have

$$\begin{aligned}C^0_{\pi^*}&(\boldsymbol{\nu}, \mathbf{n}, \mathbf{u})\\ &= -c_s + \inf_{\pi\in\Pi}\bigg\{\mathbb{E}\bigg[\sum_{t=0}^{\infty}\gamma^t\{(c_s+c_h)\mathbb{1}(n_{\pi(t),t}=0)\\ &\qquad + c(Z(\nu_{\pi(t),t}, n_{\pi(t),t}))\}\bigg]\bigg\} + \frac{\gamma q(c_q - c_s)}{1-\gamma}. \quad (30)\end{aligned}$$

The solution to the minimization problem on the right-hand side is the same as the solution to that minimization problem if the training cost is $c_s + c_h$, and the switching and quitting costs are set equal to 0. As a consequence, $M_i(n_{i,t}, \nu_{i,t}, u_{i,t}, c_h, c_s, c_q) < \hat{m}(c_h, c_s, c_q)$ if and only if $M_i(n_{i,t}, \nu_{i,t}, u_{i,t}, c_h + c_s, 0, 0) < \hat{m}(c_h + c_s, 0, 0)$ for all $t \geq 0$. To complete our argument, we then need to prove (27). The left-hand side satisfies

$$\mathbb{E}\bigg[\sum_{r=0}^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})-1} \gamma^{r+1} q\bigg] = \sum_{r=1}^{\infty} \gamma^r q\, \mathbb{P}(\tilde{\Lambda}(\hat{\nu}, 0, 1, \hat{m}) \geq r), \quad (31)$$

and that the right-hand side satisfies

$$\begin{aligned}\mathbb{E}[\gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}\mathbb{1}_{Q_0}] &= \mathbb{E}[\gamma^{\tilde{\Lambda}(\hat{\nu},0,1,\hat{m})}\mathbb{1}(\tilde{\Lambda}(\hat{\nu}, 0, 1, \hat{m}) = \Lambda)]\\ &= \sum_{r=1}^{\infty}\gamma^r\, \mathbb{P}(\Lambda = r, \tilde{\Lambda}(\hat{\nu}, 0, 1, \hat{m}) = r). \quad (32)\end{aligned}$$

By using the shorthand $\tilde{\Lambda} \equiv \tilde{\Lambda}(\hat{\nu}, 0, 1, \hat{m})$, recalling that $\tilde{\Lambda} \stackrel{d}{=} 1 + \Lambda(\boldsymbol{\pi}^*)$, and using the definition for the quitting probability, q, in (4), we have

$$\begin{aligned}q\,\mathbb{P}(\tilde{\Lambda} \geq r) &= \mathbb{P}(L_{r-1} = 1 \mid \Lambda(\boldsymbol{\pi}^*) \geq r-1)\,\mathbb{P}(\tilde{\Lambda} \geq r)\\ &= \mathbb{P}(L_{r-1} = 1 \mid \tilde{\Lambda} \geq r)\,\mathbb{P}(\tilde{\Lambda} \geq r)\\ &= \mathbb{P}(L_{r-1} = 1, \tilde{\Lambda} \geq r),\end{aligned}$$

where the last equality follows from the definition of conditional probability. Recall from (10) that $\mathbb{P}(L_{r-1} = 1, \tilde{\Lambda} \geq r) = \mathbb{P}(\Lambda = r, \tilde{\Lambda} \geq r)$, and because $\Lambda = r$ implies $\tilde{\Lambda} \leq r$, we also have $\mathbb{P}(\Lambda = r, \tilde{\Lambda} \geq r) = \mathbb{P}(\Lambda = r, \tilde{\Lambda} = r)$, which in turn implies that $q\,\mathbb{P}(\tilde{\Lambda} \geq r) = \mathbb{P}(\Lambda = r, \tilde{\Lambda} = r)$, just as needed in (31) and (32) to complete the proof of (27). □

References

Akşin OZ, Armony M, Mehrotra V (2007) The modern call center: A multi-disciplinary perspective on operations management research. *Production Oper. Management* 16(6):665–688.

Arlotto A, Gans N, Chick SE (2010) Optimal employee retention when inferring unknown learning curves. Johansson B, Jain S, Montoya-Torres J, Hugan J, Yücesan E, eds. *Proc. 2010 Winter Simulation Conf.* (IEEE, Piscataway, NJ), 1178–1188.

Asawa M, Teneketzis D (1996) Multi-armed bandits with switching penalties. *IEEE Trans. Automatic Control* 41(3):328–348.

Bailey CD (1989) Forgetting and the learning curve: A laboratory study. *Management Sci.* 35(3):340–352.

Banks JS, Sundaram RK (1992) Denumerable-armed bandits. *Econometrica* 60(5):1071–1096.

Banks JS, Sundaram RK (1994) Switching costs and the Gittins index. *Econometrica* 62(3):687–694.

Bergemann D, Välimäki J (2001) Stationary multi-choice bandit problems. *J. Econom. Dynam. Control* 25(10):1585–1594.

Bertsekas DP, Shreve SE (1978) *Stochastic Optimal Control*, Vol. 139 (Academic Press, New York).

Brown L, Gans N, Mandelbaum A, Sakov A, Shen H, Zeltyn S, Zhao L (2005) Statistical analysis of a telephone call center: A queueing-science perspective. *J. Amer. Statist. Assoc.* 100(469):36–50.

Carillo JE, Gaimon C (2000) Improving manufacturing performance through process change and knowledge creation. *Management Sci.* 46(2):265–288.

Dada M, Srikanth KN (1990) Monopolistic pricing and the learning curve: An algorithmic approach. *Oper. Res.* 38(4):656–666.

Farias VF, Madan R (2011) The irrevocable multiarmed bandit problem. *Oper. Res.* 59(2):383–399.

Frazier P, Powell W, Dayanik S (2009) The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput.* 21(4):599–613.

Freeman PR (1983) The secretary problem and its extensions: A review. *Internat. Statist. Rev.* 51(2):189–206.

Gaimon C (1997) Planning information technology-knowledge worker systems. *Management Sci.* 43(9):1308–1328.

Gaimon C, Ozkan GF, Napoleon K (2011) Dynamic resource capabilities: Managing workforce knowledge with a technology upgrade. *Organ. Sci.* 22(6):1560–1578.

Gans N, Zhou Y-P (2002) Managing learning and turnover in employee staffing. *Oper. Res.* 50(6):991–1006.

Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* 5(2):79–141.

Gans N, Liu N, Mandelbaum A, Shen H, Ye H (2010) Service times in call centers: Agent heterogeneity and learning with some operational consequences. Berger JO, Cai TT, Johnstone IM, eds. *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown*, Vol. 6 (Institute of Mathematical Statistics, Beachwood, OH), 99–123.

Gittins JC (1979) Bandit processes and dynamic allocation indices. *J. Roy. Statist. Soc. Ser. B* 41(2):148–177.

Goldberg MS, Touw AE (2003) Statistical methods for learning curves and cost analysis. *Topics in Operations Research* (INFORMS, Catonsville, MD).

Huselid MA (1995) The impact of human resource practices on turnover, productivity, and corporate financial performance. *Acad. Management J.* 38(3):635–672.

Jovanovic B (1979) Job matching and the theory of turnover. *J. Political Econom.* 87(5):972–990.

Jun T (2004) A survey on the bandit problem with switching costs. *De Economist* 152(4):513–541.

Lapré MA, Mukherjee AS, Van Wassenhove LN (2000) Behind the learning curve: Linking learning activities to waste reduction. *Management Sci.* 46(5):597–611.

March JG (1991) Exploration and exploitation in organizational learning. *Organ. Sci.* 2(1):71–87.

Mazzola JB, McCardle KF (1996) A Bayesian approach to managing learning-curve uncertainty. *Management Sci.* 42(5):680–692.

Mazzola JB, McCardle KF (1997) The stochastic learning curve: Optimal production in the presence of learning-curve uncertainty. *Oper. Res.* 45(3):440–450.

Müller A, Stoyan D (2002) *Comparison Methods for Stochastic Models and Risks*, Wiley Series in Probability and Statistics (John Wiley & Sons, Chichester, UK).

Nagypál É (2007) Learning by doing vs. learning about match quality: Can we tell them apart? *Rev. Econom. Stud.* 74(2): 537–566.

Nembhard DA (2001) Heuristic approach for assigning workers to tasks based on individual learning rates. *Internat. J. Production Res.* 39(9):1955–1968.

Nembhard DA, Osothsilp N (2002) Task complexity effects on between-individual learning/forgetting variability. *Internat. J. Indust. Ergonomics* 29(5):297–306.

Nembhard DA, Uzumeri MV (2000a) Experiential learning and forgetting for manual and cognitive tasks. *Internat. J. Indust. Ergonomics* 25(4):315–326.

Nembhard DA, Uzumeri MV (2000b) An individual-based description of learning within an organization. *IEEE Trans. Engrg. Management* 47(3):370–378.

Niño-Mora J (2008) A faster index algorithm and a computational study for bandits with switching costs. *INFORMS J. Comput.* 20(2):255–269.

Pinker EJ, Shumsky RA (2000) The efficiency-quality trade-off of cross-trained workers. *Manufacturing Service Oper. Management* 2(1):32–48.

Pisano GP, Bohmer RMJ, Edmondson AC (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci.* 47(6):752–768.

Shafer SM, Nembhard DA, Uzumeri MV (2001) The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Sci.* 47(12):1639–1653.

Shaked M, Shanthikumar JG (2007) *Stochastic Orders*, Springer Series in Statistics (Springer, New York).

Shen H (2003) Nonparametric regression for problems involving lognormal distributions. Ph.D. dissertation, University of Pennsylvania, Philadelphia. http://search.proquest.com/docview/305307164.

Shen H, Brown LD (2006) Non-parametric modelling for time-varying customer service time at a bank call centre. *Appl. Stochastic Models Bus. Ind.* 22(3):297–311.

Sundaram RK (2005) Generalized bandit problems. Austen-Smith D, Duggan J, eds. *Social Choice and Strategic Decisions—Essays in Honor of Jeffrey S. Banks*, Studies in Choice and Welfare (Springer-Verlag, Berlin), 131–162.

Uzumeri M, Nembhard D (1998) A population of learners: A new way to measure organizational learning. *J. Oper. Management* 16(5):515–528.

Weitzman ML (1979) Optimal search for the best alternative. *Econometrica* 47(3):641–654.

Whitt W (2006) The impact of increased employee retention on performance in a customer contact center. *Manufacturing Service Oper. Management* 8(3):235–252.

Whittle P (1980) Multi-armed bandits and the Gittins index. *J. Roy. Statist. Soc. Ser. B* 42(2):143–149.

Wiersma E (2007) Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Sci.* 53(12):1903–1915.

Yelle LE (1979) The learning curve: Historical review and comprehensive survey. *Decision Sci.* 10(2):302–328.