Optimal Hiring and Retention Policies for Heterogeneous Workers who Learn

Alessandro Arlotto

OPIM Department; The Wharton School; Univ. of Pennsylvania; Philadelphia, PA, 19104-6340, U.S.A. alear@wharton.upenn.edu

Stephen E. Chick

INSEAD; Technology & Operations Management Area; Blvd. de Constance; 77305 Fontainebleau France; stephen.chick@insead.edu

Noah Gans

OPIM Department; The Wharton School; Univ. of Pennsylvania; Philadelphia, PA, 19104-6340, U.S.A. gans@wharton.upenn.edu

On-the-job learning is an important element of the operations of call-centers, manufacturing and other activities, especially when there may be high turnover of employees or of new technologies and processes. Workers may have diverse capabilities that change through time, and different policies for monitoring and hiring will influence the long-run performance of a firm. In this paper, we study the hiring and retention of heterogeneous workers who learn over time. We formulate the problem as an infinite-armed bandit and characterize the optimal hiring and retention policy in detail. We develop approximations that allow the efficient implementation of the optimal policy and the evaluation of its performance. Our numerical examples demonstrate under different scenarios that the value of an active monitoring and screening of employees can be substantial.

1. Introduction

Workers are heterogeneous, and they evolve over time. Evolution often takes the form of on-the-job learning, with attendant decreases in the time required to complete tasks or improvements in quality. When employees turn over they may be replaced by new hires who differ in ability and experience.

Often there may be uncertainty regarding employee attributes. 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 employees, across tasks, and over time – makes decisions regarding the retention of workers complex. The longer a worker is retained, the better an inference an employer can make regarding his or her attributes. On-the-job learning, which can lead to quality improvements in incumbent employees, also favors employee retention. Yet 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 he wants to continue working or not. Given the worker decides to stay, the employer can decide whether to retain him or to replace him with a new hire.

In Section 3, we formulate this problem as an infinite-horizon, discounted problem in which, at any time, the employer uses a single worker. In Section 4, we show that it is, essentially, a multi-armed bandit problem with an infinite number of arms. Banks and Sundaram (1992) analyzed a version of this problem in which a given arm (employee) can be one of a finite number of types. In our Bayesian setting, arm types correspond to a continuum of prior distributions with compact support and are therefore more complex. Nonetheless, we prove that several of their core results hold in our case as well:

- The employer can use a worker's prior distribution and tenure to calculate a so-called Gittins index, and at any time it is optimal for the employer to use a Gittins-index minimal employee.
- It is optimal to retain current employees as long as their Gittins indices compare favorably to those of potential hires.
- If a current employee's Gittins index is not minimal, however, then it is optimal to hire a new worker and to never return to the current employee.

This last property is known as "no-recall" and is particularly interesting from an application perspective. For recent developments concerning bandit problems that do not recall, or equivalently that are irrevocable, we refer the reader to Farias and Madan (2011).

In Section 5, we indicate how these Gittins-index results extend to more complex settings: those in which the employer retains multiple employees as well as those in which she hires from multiple, heterogeneous pools of potential hires. In both cases our original results regarding "no-recall" properties generalize directly.

Given the availability of a Gittins index, the above policy is both intuitive and straightforward to execute. Unfortunately, the Gittins index itself is difficult to calculate. Nevertheless, in Section 6, for specific common forms of the learning-curve function, and given performance that can be appropriately transformed into normally-distributed data with known sampling variances and unknown means (a conjugate prior distribution), we:

- show that, for a fixed level of experience, the Gittins index is monotone in the posterior mean of the unknown parameter, which allows us to delineate a simple stopping boundary, below which a current worker's employment should continue and above which it should stop;

- develop approximations to the Gittins index that are straightforward to calculate and implement.

The approximations developed in Section 6 are the basis for numerical examples in Section 7 that provide insights into the economic nature of the hiring and retention problem. In particular, we:

- demonstrate that the stopping boundary reflects a tradeoff between two types of learning: the employee's performance improvement that is linked to on-the-job experience, and the employer's statistical learning that allows for better judgment concerning the worker's ability;

- show that the value of active monitoring and screening of employees can be substantial;

- observe that the early stages of workers' tenure are the most important for the effectiveness of the optimal Gittins-index policy;

- suggest that simple hiring policies with a trial period followed by a one-shot hiring and retention decision have the potential to perform well, within a few percent of the optimal Gittins-index policy.

Sensitivity analysis with respect to model parameters provides additional insights:

- in addition to direct gains that accrue from the 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.

2. Literature review

There is a vast empirical literature on learning-curve phenomena at the individual and organizational levels (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 level (e.g., Nembhard and Uzumeri 2000a, Nembhard 2001) and organizational level (e.g., Bailey 1989, Lapré et al. 2000, Pisano et al. 2001). Nembhard and Uzumeri (2000b) provide a unified study that considers both of them. Our analysis focuses on the individual level.

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 multi-armed bandit problems that addresses matching problems in labor-market: typically, problems in which employees choose firms (Jovanovic 1979, Banks and Sundaram 1992). In our context this work can be reinterpreted as addressing firms choosing employees.

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. But that paper's aims and results differ significantly from ours. While its model and analysis enable the use of statistical methods to discriminate between the two forms of learning in empirical employment records, they do not provide the insights into the nature of effective retention/termination decisions that are the focus of our work.

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 considers 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) study how individuals respond to production changes. They analyze the relationship between task complexity and the distribution of individual learning and forgetting parameters. Their main conclusion is that task complexity affects these measures. Gans et al. (2010) show that the service times of call-center agents reflect on-the-job learning, as well as agent heterogeneity.

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 gives 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 Aksin 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, which may encompass the call center, manufacturing and hospital examples mentioned above. Gaimon et al. (to appear) use mathematical models and empirical data to assess learning-before-doing, which can be modeled as hiring 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.

The technical aspects of the paper rely on related literature regarding infinite-armed bandit models. Berry and Fristedt (1985) and Gittins (1989) are standard references. Easley and Kiefer (1988), Banks and Sundaram (1992) and Bergemann and Välimäki (2001) provide results that are of use in our analysis below.

3.   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 so that workers who quit or are terminated need not be considered again. Section 5 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 and is defined by the relation

$$Z_{i,n_i} = g(\boldsymbol{\theta}_i, n_i, \epsilon_{i,n_i}), \text{ for all } i \in \mathcal{S}_0, \tag{1}$$

where $\boldsymbol{\theta}_i \in \Omega$ is a vector of parameters that reflects worker i's ability, $n_i = 0, 1, 2, \ldots$ reflects his experience, ϵ_{i,n_i} 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,n_i} by z_{i,n_i}.

For example, for $\boldsymbol{\theta}_i = (a_i, b_i)$, Yelle (1979) describes the following commonly-used form:

$$Z_{i,n_i} = \exp\left(a_i + b_i \ln(n_i + 1) + \epsilon_{i,n_i}\right), \quad n_i = 0, 1, 2, \ldots. \tag{2}$$

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,n_i} 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 Section 4, require only the general functional form (1), together with some technical assumptions. While our analysis does hinge on a single measure of performance, the representation of an outcome, Z_{i,n_i}, can be generalized to explicitly represent multiple dimensions (such as revenue, cost, quality) and then aggregated into a single score by using a functional. Section 6, in which we develop methods for explicitly calculating the stopping boundaries necessary to implement optimal policies, assumes a more specific form of Z_{i,n_i}, 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, $\mathbf{L}_i = (L_{i,0}, L_{i,1}, L_{i,2}, \ldots)$ such that worker i leaves or quits *after* his $(n_i + 1)$st performance if and only if $L_{i,0} = L_{i,1} = \cdots = L_{i,n_i-1} = 0$ and $L_{i,n_i} = 1$. We denote the realization of $\mathbf{L}_i$ and L_{i,n_i} by $\boldsymbol{\ell}_i$ and ℓ_{i,n_i} respectively.

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 until he quits or is terminated. In turn, we define worker i's *quitting* probability, q_{i,n_i}, to be

$$q_{i,n_i} = \mathbb{P}\left(L_{i,n_i} = 1 | \Lambda_i(\pi) \geq n_i + 1\right), \tag{4}$$

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

Let $\mathcal{H}_{i,n_i} = \{(z_{i,s}, \ell_{i,s}) : 0 \leq s \leq n_i - 1\}$ denote worker i's *employment history* when his experience is n_i. The quitting probability of an employee with experience n_i, q_{i,n_i}, may depend on $\mathcal{H}_{i,n_i}$ and on his ability $\boldsymbol{\theta}_i$, but it must be independent of the employer's hiring policy, π:

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

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 patters 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, however, 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 Section 7.

The employer does not know each employee's $\boldsymbol{\theta}_i$ or $\boldsymbol{\ell}_i$ in advance. Rather, she believes that there exists a random vector, $\boldsymbol{\Theta}$, that represents ability 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 *iid* samples from the population distributions. At time $t = 0$ all potential workers, i, have the same history, $\mathcal{H}_{i,0} = \emptyset$, and the same prior distribution for $\boldsymbol{\Theta}_i$, so the employer is indifferent among her choices. For each worker, i, and for each level of cumulative experience, n_i, the employer uses i's employment history, $\mathcal{H}_{i,n_i}$, to update her beliefs concerning the distribution of the parameter $\boldsymbol{\Theta}_i$.

For any Borel set $X \subseteq \Omega$ we let $\nu_{i,n_i}(X) = \mathbb{P}(\boldsymbol{\Theta}_i \in X | \mathcal{H}_{i,n_i})$ denote the posterior probability distribution that describes the employer's uncertainty concerning $\boldsymbol{\Theta}_i$ *after* the n_ith performance. For $\boldsymbol{\Theta}_i \sim \nu_{i,n_i}$ we let $Z_{i,n_i} \equiv Z(\nu_{i,n_i}, n_i) \equiv Z(\nu_i, n_i)$, and for $\{\boldsymbol{\Theta}_i = \boldsymbol{\theta}_i\}$ we assume that worker i's performance $\{Z(\nu_{i,n_i}, n_i) | \boldsymbol{\theta}_i\}$ has density $\xi_{n_i}(z | \boldsymbol{\theta}_i)$.

An untried worker, i, has prior experience $n_i \equiv 0$ and prior distribution $\nu_{i,0} \equiv \nu_0$, where ν_0 is a common prior distribution on $\boldsymbol{\Theta}_i$ for all potential workers. After the completion of a task, worker i's experience,

n_i, increases deterministically by one, and the employer updates the distribution of her belief concerning i's ability according to Bayes' rule. If $\mathcal{P}(\Omega)$ is the set of all probability measures, ν, on Ω, then the Bayes operator $\beta : \mathcal{P}(\Omega) \times \mathbb{R} \to \mathcal{P}(\Omega)$ is defined as

$$\mathbb{P}\left(\boldsymbol{\Theta}_i \in X | \mathcal{H}_{i,n_i}\right) = \nu_{i,n_i+1}(X) = \beta(\nu_{i,n_i}, z)(X) = \frac{\int_X \xi_{n_i}(z|\boldsymbol{\theta})\, d\nu_{i,n_i}}{\int_\Omega \xi_{n_i}(z|\boldsymbol{\theta})\, d\nu_{i,n_i}}, \tag{5}$$

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

The employer's costs are driven by both employer and employee actions. If, at the start of a period, the employer hires a new employee, she incurs hiring cost, c_h, that includes expected recruiting and training costs. If, at the end of a period, the employee quits, the employer bears a quitting cost, c_q, that includes potential separation costs. If the employee does not quit, then the employer may decide to terminate him, in which case she bears an analogous firing cost, c_f.

Within each period, the employer incurs a task-related cost that is driven by the selected employee's performance, $c(z_{i,n_i})$. We assume that $c(z)$ is continuous and increasing 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, ν_{i,n_i}, to estimate his expected task-related cost:

$$\mathbb{E}[c\left(Z(\nu_{i,n_i}, n_i)\right)] = \int_\Omega \left(\int_\mathcal{E} c\left(g(\boldsymbol{\theta}, n_i, s)\right) \xi_{n_i}(g(\boldsymbol{\theta}, n_i, s)|\boldsymbol{\theta})\, ds\right)\, d\nu_{i,n_i}. \tag{6}$$

The employer discounts all of these costs at rate $\gamma \in (0,1)$ and for any hiring and retention policy, π, her expected employment costs over an infinite horizon are

$$\begin{aligned}
C_\pi(\nu_0) = c_h + \mathbb{E}\Bigg[\sum_{t=0}^\infty \gamma^t \Bigg\{ &\ c\left(Z(\nu_{\pi(t)}, n_{\pi(t)})\right) \\
&+ (c_f + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1), \pi(t-1) \in \mathcal{S}_t) \\
&+ (c_q + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1), \pi(t-1) \notin \mathcal{S}_t)\ \Bigg\}\Bigg].
\end{aligned} \tag{7}$$

The quantity c_h outside the expected value is the hiring cost of the first employee, and the terms within the summation reflect the three possible sources of cost in each period, t. The first term, $c\left(Z(\nu_{\pi(t)}, n_{\pi(t)})\right)$, reflects employee $\pi(t)$'s task-related costs. The second term, $(c_f + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1), \pi(t-1) \in \mathcal{S}_t)$, is the cost of hiring a new worker at time t, should the previous employee be terminated. The third term, $(c_q + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1), \pi(t-1) \notin \mathcal{S}_t)$, reflects the cost of hiring a new worker at time t, should the previous employee quit.

When selecting which employee to utilize at time t, the employer uses only the prior distribution ν_0, the cumulative experience prior to time t of each worker, i, $n_{i,t} = \sum_{s=0}^{t-1} \mathbb{1}(\pi(s) = i)$, and the employment histories of eligible employees, $\mathcal{H}_t = \cup_{i \in \mathcal{S}_t}\{\mathcal{H}_{i,n_{i,t}}\}$. More formally, we let Π denote the set of *non-anticipating*

hiring policies, and we assume that the employer seeks a policy $\pi^* \in \Pi$ that minimizes the expected discounted value of future employment costs

$$\pi^* = \arg\min_{\pi \in \Pi} C_\pi(\nu_0). \tag{8}$$

For the problem to be analytically tractable we assume that the parameter space Ω is a compact subset of $\mathbb{R}^d$ and that ν_0 has support Ω. We require that, for each n, $g(\boldsymbol{\theta}, n, u)$ is jointly continuous in $(\boldsymbol{\theta}, u)$ and uniformly bounded, so that $g(\boldsymbol{\theta}, n, u) \in [K_{\inf}, K_{\sup}]$ for each triple $(\boldsymbol{\theta}, n, u)$. Further, we assume that, for each n, the marginal density $\xi_n(z|\boldsymbol{\theta})$ is jointly continuous in $(\boldsymbol{\theta}, z)$, with support $[K_{\inf}, K_{\sup}]$. (See, e.g. Easley and Kiefer 1988.)

4. Structure of the Optimal Policy

The employer's hiring and retention problem is closely related to classic Bayesian bandit problems. Each potential employee's state is a triple that represents his experience, n_i, the employer's current belief regarding his ability, ν_{i,n_i} and an availability index r_i such that $r_i = 1$ if worker i has quit and $r_i = 0$ otherwise. In each period, the employer chooses exactly one employee, and this choice yields a single-period cost, along with a state transition of only that employee; the states of all other potential hires remain unchanged. From (4) – (5) we see that state transitions are Markov.

One small difference between the employer's problem and traditional bandit problems is that, in the former, an employee becomes unavailable when he quits, while in the latter arms are always available. Is it not difficult, however, to transform the hiring and retention problem so that it neatly fits within the bandit framework. To do so, we modify the ability of a worker who quits so that he would never be hired by the employer. More specifically, we modify (5) as follows. We choose some $K \in (K_{\sup}, \infty)$ such that $c(K_{\sup}) + c_h + \max\{c_f, c_q\} < c(K)$ and let

$$\nu_{i,n_i+1}(X) = \begin{cases} \beta(\nu_{i,n_i}, z)(X) & \text{if } \max\{\ell_{i,k}, \ 0 \le k \le n_i\} = 0 \\ \mathbb{1}_K & \text{if } \max\{\ell_{i,k}, \ 0 \le k \le n_i\} = 1, \end{cases} \tag{9}$$

where $X \subseteq \Omega$, β is the Bayes operator defined in (5), and $Z(\mathbb{1}_K, n) = K$ for every n. Thus, rather than quitting at n_i, employee i becomes unproductive, and his cost exceeds the cost of any possible realization of any worker who has not yet quit, plus the largest cost of hiring a new worker. Then

$$\Lambda_i = \inf\{n \in \mathbb{N} : \ Z(\nu_{i,n}, n) = K\} \tag{10}$$

is the time at which he becomes unproductive. Note that, for each policy $\pi \in \Pi$, if employee i quits before he is terminated, then $1 + \Lambda_i(\pi) = \Lambda_i$. Otherwise, $1 + \Lambda_i(\pi) < \Lambda_i$.

Call the original problem in (8), 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 effect of workers who become unproductive is analogous to that of those who quit.

LEMMA 1.　(i) *In Problem 2, any policy that uses worker i when $n_i \geq \Lambda_i$ is never optimal.*

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

Proofs of these claims and of others below are found in Appendix.

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

Problem 2 is a variant of the infinite-arm bandit problem analyzed by Banks and Sundaram (1992). Their paper's core results can be extended to our problem setting, and we use them characterize the structure of the optimal hiring and retention policy. More specifically, we derive the optimal policy by solving a family of stopping problems in which, at each period, t, the employer chooses between employing a single worker, $i \in \mathcal{S}_t$, or terminating all employment and paying the firing cost, c_f, and 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 experience index, n, 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) to study denumerable-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 1996, Prop. 9.8).[1] The optimal value function for the retirement-option approach satisfies the following Bellman equation:

$$V(\nu, m, n, r) = \begin{cases} \min\{c_f + m, HV(\nu, m, n)\}, & \text{if } r = 0, \\ \min\{c_q + m, c(K) + \gamma V(\mathbb{1}_K, m, n+1, 1)\}, & \text{if } r = 1, \end{cases} \quad (11)$$

where

$$HV(\nu, m, n) = c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))] + \gamma(1 - q_n)\mathbb{E}[V(\beta(\nu, Z(\nu, n)), m, n+1, 0)] \quad (12)$$

$$+ \gamma q_n V(\mathbb{1}_K, m, n+1, 1),$$

In words, when the current employee is productive ($r = 0$), the employer has the choice of terminating the employee at cost c_f and incurring the retirement payment, m, or of continuing his employment. Here, the expected discounted value of continuing, HV, is the expected cost of the current period plus the expected discounted cost to go. The expected discounted cost to go is determined by conditioning on whether or not the employee remains productive (which happens with probability $1 - q_n$). When the current employee has

---

[1] A policy is *stationary* if, at any time t, the action it prescribes in a given state is independent of t. A policy is *deterministic* if the action it prescribes is never randomized.

become unproductive ($r = 1$), then the employer has the choice between incurring the quitting cost, c_q, and the retirement payment, m, or continuing his employment. In this case, the expected discounted value of continuing employment is the sum of the current period cost of an unproductive worker, $c(K)$, plus the associated cost to go, which are both deterministic.

When $c_q + m \leq c(K)/(1 - \gamma)$, so that retiring is attractive when $r = 1$, (12) becomes

$$HV(\nu, m, n) = c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))] + \gamma(1 - q_n)\mathbb{E}[V(\beta(\nu, Z(\nu, n)), m, n + 1, 0)] + \gamma q_n(c_q + m).$$

The last addend represents the cost paid for an employee who has quit, c_q, plus the retirement cost for the employer, m.

Given the availability of the value function (11) we are interested in the value of m for which the employer is indifferent between continuing to employ the current hire or terminating him (at cost c_f if $r = 0$ or c_q if $r = 1$) and retiring (at cost m). We denote that value by the index

$$M(\nu, n, r) = \begin{cases} \sup\{m \in \mathbb{R} : V(\nu, m, n, 0) = c_f + m\} & \text{if } r = 0 \\ \sup\{m \in \mathbb{R} : V(\nu, m, n, 1) = c_q + m\} & \text{if } r = 1. \end{cases} \tag{13}$$

This index is well-defined because the value function (11) is concave and non-decreasing in m, a fact that is stated and proved in the appendix.

Theorem 1 shows that there exists an optimal hiring and retention policy that always selects an employee with a minimal index (13). The index is therefore justifiably called a *Gittins index* for this problem. With it, we can characterize the optimal hiring and retention policy (8).

THEOREM 1. *Worker i with experience n_i and availability index r_i is an optimal selection at time t if and only if $M_i(\nu_{i,n_i}, n_i, r_i) = \inf_{j \in \mathcal{S}_t}\{M_j(\nu_{j,n_j}, n_j, r_j)\}$. Furthermore, there exists an optimal policy that always selects a worker with a minimal Gittins index.*

The optimal policy described in Theorem 1 implies that there is often just one Gittins-index-minimal employee.

COROLLARY 1. *Let $m_0 = M(\nu_0, 0, 0)$ be the Gittins index of a worker who has not yet been tried. If at $t = 0$ all potential hires have Gittins indices of m_0, then at any time, t, at most one worker, i, has Gittins index $M_i(\nu_{i,n_i}, n_i, 0) < m_0$.*

Together Lemma 1 and Theorem 1 also imply the following useful "no-recall" property.

COROLLARY 2 (**"No-Recall" Property**). *Let π^* be an optimal Gittins-index policy and let $t_i = \inf\{t : \pi^*(t) = i\}$ be the first time worker i is employed. Then:*

(i) *Worker i is employed continuously for $\Lambda_i(\pi^*)$ periods; that is $\pi^*(t) = i$ for all $t_i \leq t < t_i + \Lambda_i(\pi^*)$.*

(ii) *It is never optimal to employ worker i from time $t_i + \Lambda_i(\pi^*)$ on; that is $\pi^*(t) \neq i$ for all $t_i + \Lambda_i(\pi^*) \leq t$.*

Therefore, it is never optimal to employ a worker who was previously terminated, and the optimal hiring policy takes the following form. At the start of the problem, the employer calculates the index for untried workers, m_0. Then she chooses an employee, i, at random from the pool of untried employees, and after each performance, she recalculates his Gittins index based on the posterior distribution ν_{i,n_i}. If the new Gittins index has a value of m_0 or less, then it is optimal to retain the current employee. If the updated Gittins index rises above m_0 then it is optimal to terminate him and hire a new employee, at random, from the pool.

For an employer seeking to retain a single employee, the hiring and retention problem 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 m_0, whichever comes first.

Given the *iid* nature of the stopping problems, we can show that the Gittins index of the untried workers is closely related to the total expected discounted cost under the optimal policy.

THEOREM 2. *Let π^* be a Gittins-index policy, and let $m_0 = M(\nu_0, 0, 0)$ be the Gittins index of the untried workers. If $\mathbb{E}[\Lambda_1] < \infty$ then $m_0 + c_f(1 - \gamma\mathbb{E}\left[\gamma^{\Lambda_1(\pi^*)}\right])^{-1} = \inf_{\pi \in \Pi} C_\pi(\nu_0)$.*

Theorem 2 is appealing since it links the expected total discounted cost under the optimal policy to the Gittins index. This type of result does not usually hold in a general bandit setting. Here, it relies on the "no-recall" property of the optimal policy described in Corollary 2. This allows us to interpret our hiring process as a discounted renewal reward process in which the tenure of every worker is the length of the renewal interval, and the cost of each worker throughout his tenure is the reward. Due to the "no-recall" property, the renewal intervals as well as the rewards are *iid*. In Section 6, we use Theorem 2 to estimate the expected discounted value of implementing a Gittins-index policy.

5. Extensions: Multiple Parallel Workers and Different Pools

Sections 3 and 4 considered the problem of employing a single worker. We now consider two extensions. Section 5.1 considers an employer who wishes to retain multiple employees who work in parallel. Section 5.2 considers the problem in which distinct (infinite) pools of heterogeneous workers are available. In both cases, the optimality of an index rule is retained.

5.1. Hiring and Retention of Multiple Workers

If the employer wishes to retain a fixed number, D, of people working in parallel, a Gittins-index policy remains optimal. In fact, one can partition the infinite pool of potential employees, $\mathcal{S}_t$, into D separate, countably infinite pools, $\mathcal{S}_{1,t}, \mathcal{S}_{2,t}, \ldots, \mathcal{S}_{D,t}$, in which untried workers have common prior distribution, ν_0. Further, when employee i in pool d quits at time t, he is removed from that pool so that $\mathcal{S}_{d,t+1} = \mathcal{S}_{d,t} \setminus \{i\}$.

The infinite-horizon total expected discounted cost for the employer is then

$$
C_{\boldsymbol{\pi}}^D(\nu_0) = c_h D + \mathbb{E}\Bigg[\sum_{t=0}^{\infty} \gamma^t \sum_{d=1}^{D} \bigg(c(Z(\pi_d(t), n_{\pi_{d}(t)})) \tag{14}
$$
$$
+ (c_f + c_h)\mathbb{1}(\pi_d(t) \neq \pi_d(t-1), \pi_d(t-1) \in \mathcal{S}_{d,t})
$$
$$
+ (c_q + c_h)\mathbb{1}(\pi_d(t) \neq \pi_d(t-1), \pi_d(t-1) \notin \mathcal{S}_{d,t}) \bigg) \Bigg],
$$

where $\pi_d(t) \in \mathcal{S}_{d,t}$ identifies the index of the worker who is employed from pool d at time t, and $n_{\pi_d(t)}$ represents his experience.

By interchanging the sums in (14) one obtains $C_{\boldsymbol{\pi}}^D(\nu_0) = \sum_{d=1}^{D} C_{\pi}^d(\nu_0)$, where $C_{\pi}^d(\nu_0)$ 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. Hence, at any time, t, in 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, due to Bergemann and Välimäki (2001), crucially depends on the assumption that all the workers have the same prior distribution, ν_0, 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 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).

5.2. 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 Theorem 1.

For $t = 0, 1, 2, \ldots$ consider, without loss of generality, two infinite pools $\mathcal{S}_t'$ and $\mathcal{S}_t''$, where the untried workers have common prior distributions ν_0' and ν_0'', with $\nu_0' \neq \nu_0''$.

Let $M(\nu_0', 0, 0)$ and $M(\nu_0'', 0, 0)$ be the indices of the untried workers in each pool. If $M(\nu_0', 0, 0) \neq M(\nu_0'', 0, 0)$, then workers belonging to the pool with larger index are never employed. Otherwise, if $M(\nu_0', 0, 0) = M(\nu_0'', 0, 0)$, then the employer is indifferent between the two populations.

6. Implementing the Optimal Policy

This section shows how analytic properties of the hiring and retention problem can be combined with dynamic programming techniques to enable the computation of the relevant Gittins indices when performance has certain structural properties. Because our set of *iid* stopping problems allows us to focus on a single employee we drop the subscript, i, hereafter for clarity.

As shown in the appendix, for any given ν, n and r the value function, $V(\nu, n, m, r)$, is concave and nondecreasing in m. Therefore, given ν, n, and r a simple search scheme, such as bisection, can be used to find the largest fixed point, $M(\nu, n, r)$, that defines the Gittins index.

We now turn to the computation of $V(\nu, n, m, r)$. We explicitly define the functional form of the $(n+1)$st performance for a worker, Z_n, in (1) in order to calculate solution values. In particular, we assume that $g(\cdot)$ is invertible and that

$$g^{-1}(Z_n) = A + h(n) + \epsilon_n, \quad n = 0, 1, 2, \ldots, \tag{15}$$

is a linear regression 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 standard deviation σ. Since 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, ν_0, that is normally distributed with mean μ_0 and variance σ_0^2, $N(\mu_0, \sigma_0^2)$.

The form in (15) implies another structural property that will be useful for computing the Gittins indices of workers. The random variables $X_n = g^{-1}(Z_n) - h(n)$ are normally distributed with unknown mean A and variance $\sigma^2 + \sigma_0^2$. By standard Bayesian analysis, ν_n, the posterior distribution of A after observing n tasks, $\mathbf{z}_n = (z_0, z_2, \ldots, z_{n-1})$, is normal with

$$\mathbb{E}[A \mid \mathbf{z}_n] = \frac{\mu_0 \frac{\sigma^2}{\sigma_0^2} + \sum_{k=0}^{n-1}(g^{-1}(z_k) - h(k))}{n + \sigma^2/\sigma_0^2} \text{ and } \mathrm{Var}[A \mid \mathbf{z}_n] = \frac{\sigma_0^2 \sigma^2}{\sigma^2 + n\sigma_0^2}.$$

Define $s_0 = \sigma^2/\sigma_0^2$, and let $s = s_0 + n$, where n is the number of samples observed for the single-worker problem. Set $y_s = \mu_0 s_0 + \sum_{k=0}^{n-1}(g^{-1}(z_k) - h(k))$ and $w_s = y_s/s$. The posterior distribution, ν_n, of A given $\mathbf{z}_n$ is thus $N(w_s, \sigma^2/s)$. We can therefore describe (ν_n, n) by (w_s, s).

These assumptions are sufficient to guarantee that both the Bellman equation (11) and the Gittins index (13) are monotone in the posterior mean of A, w_s.

PROPOSITION 1. *For any fixed, m, n, and r, $V(\nu, n, m, r)$ is nondecreasing in w_s, and for any fixed n and r, $M(\nu, n, r)$ is nondecreasing in w_s.*

The monotonicity of the Gittins index with respect to w_s allows us to concisely describe the optimal policy. For each $s = s_0 + n$, there is a simple "stopping" boundary, $\mathfrak{b}(s)$, such that it is optimal to retain the employee (continue) if $w_s < \mathfrak{b}(s)$ and to terminate the employee (stop) if $w_s > \mathfrak{b}(s)$.

Arlotto et al. (2010) provides more detail for how to use the above results to approximate V and the stopping boundary, $\mathfrak{b}$, when (15) 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 $N(w_{s_0}, \sigma^2/s_0)$, so that Proposition 1 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, s) coordinate system, estimate the terminal conditions (the period in which the dynamic programming backwards 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 Section 7 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 can be written to emphasize the dependence of Z on the posterior distribution:

$$\ln(Z_n) = A + b \ln(n + 1) + \epsilon_n, \quad n = 0, 1, 2, \ldots, \tag{16}$$

where $\epsilon_n \sim N(0, \sigma^2)$. Here, (16) is consistent with empirical studies of various industries. For example, Brown et al. (2005), Shen (2003), 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 $h(\cdot)$ as well. For example, we also tested $h(n) = b \ln(1 + n/(n + \zeta))$ and obtained qualitative conclusions that are analogous to what we describe below in Section 7. Similarly, we can define a as a common, known parameter and $g^{-1}(Z_n) = a + B h(n) + \epsilon_n$ to model pools of workers with a common base level of quality and heterogeneous rate of learning. While the theoretical results described in Section 4 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 1, may not hold.

## 7. Numerical Examples and the Value of Screening

In this section, we use the methods described in Section 6 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 demonstrate that an active hiring and retention policy reduces costs and improves the pool of workers who are employed. Moreover, we perform a sensitivity analysis with respect to the key parameters of our model, and we conclude that increases in employee learning 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.

7.1. Balancing Uncertainty and Learning Effects

The first example is loosely motivated by a call center. Each $Z_{i,n}$ represents the average duration (in minutes) of the calls that agent i handles after n days of experience. We use the log-linear learning curve model (16). The distribution of the base-level performance parameter, A, has mean $\mu_0 = 0.90$ and variance $\sigma_0^2 = 0.16$, and the sampling variance in the daily average of the service times is $\sigma^2 = 0.64$. This implies an expected service time of untried agents of $\mathbb{E}[Z_{i,0}] = 3.67$. The quitting probability q_n is constant over time and $q_n = 0.01$ for all n. 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 hiring cost is $c_h = 30$ which corresponds to the expected cost of employing untried workers for approximately 10 days (2 weeks). Termination and quitting costs are set equal to 0. (See Theorem 3, below, in Section 7.3.) 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 $\mathbb{E}[Z_{i,249}] = \alpha \mathbb{E}[Z_{i,0}]$. Choosing $\alpha = 0.50$, we obtain $b = -0.1255$.

Figure 1 displays the stopping boundary associated with the Gittins index for untried employees, m_0, which, in this example, equals 5,497.7. 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 1 we know that an employee whose posterior means fall below these stopping boundaries has a Gittins index below m_0 and should be retained, and one whose posterior means fall above the stopping boundary should be replaced by a new hire from the pool of untried potential employees.

Figure 1   **Stopping Boundaries for Posterior Mean of A (left) and for $\mathbb{E}[Z_n]$ (right).**
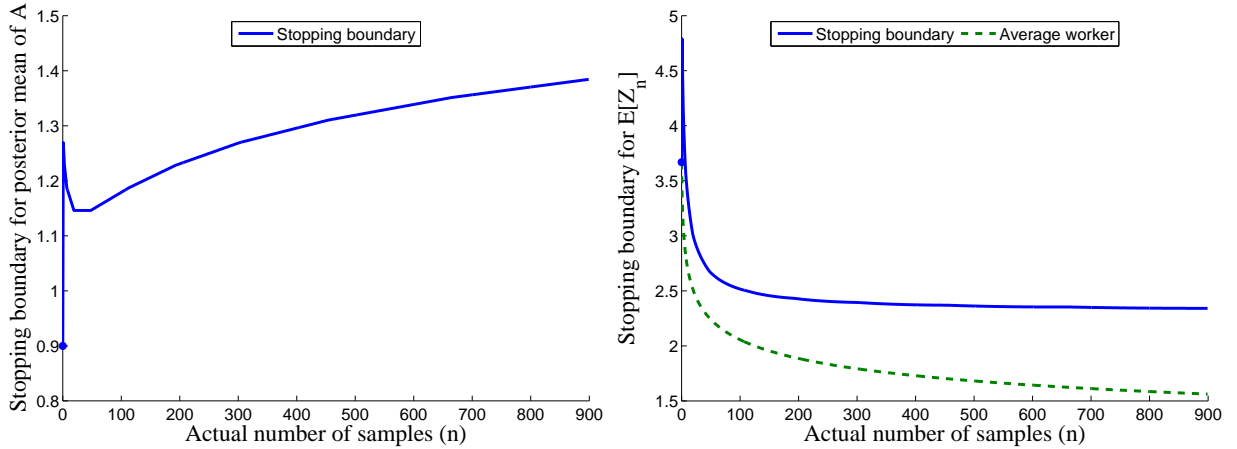
Note. Parameters: $b = -0.1255, \mu_0 = 0.90, \sigma_0 = 0.40, \sigma = 0.80, s_0 = 4, c_h = 30, c_f = c_q = 0$.

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, $\mu_0 = 0.90$, up to 1.27 is attributed to the ending of the hiring cost, c_h, which is incurred only on day zero. Afterwards, 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 right most reaches, the curve appears to increase to an asymptote involving a constant minus $h(n)$.

The right panel shows the stopping boundary with respect to the mean service time, $\mathbb{E}[Z_n]$. Here, the stopping boundary is unimodal, with a peak on day 1 due to the elimination of the day-zero hiring 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 the right panel plots $\mathbb{E}[Z_n]$ for an "average" employee with base-level service time of $A = \mu_0$. The vertical distance between the two curves is a measure of how much better or worse a "marginally retained" employee is in comparison to an "average" employee. We see that the presence of hiring 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.

Table 1   **Optimal policy and employee retention** *(standard errors for the mean in parenthesis)*.

Turnover rates	Day 1	Days 2 – 10	Days 11 – 20	Total
Terminated workers	0.0196 *(.0006)*	0.2830 *(.0020)*	0.0557 *(.0010)*	0.3982 *(.0022)*
Workers who quit	0.0102 *(.0005)*	0.0692 *(.0011)*	0.0539 *(.0010)*	0.6018 *(.0022)*

The policy terminates 39.82% of the employees: 1.96% of workers are terminated on day 1, 28.30% are terminated during periods 2 through 10, and 9.57% thereafter. Hence, much of the termination occurs early on. Of course, termination rates vary significantly with hiring costs. In Section 7.3 we present a sensitivity analysis that addresses this relationship.

7.2.   How the Optimal Policy compares with Simpler Policies

This section compares the optimal policy with four families of alternative hiring policies. In the first, workers are never terminated, and they serve until they naturally turn over.

In the second, workers are monitored using the stopping boundaries in Figure 1 for a limited screening period. During the screening period they can be terminated after each day of performance, and, 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 1, 5, 10 or 20 days long.

The third family of hiring policies 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 type of hiring policy considers 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.

Table 2 reports infinite-horizon total expected discounted costs, termination rates and long-run average service rates for each policy. The first two quantities are familiar, and we now formally define the third. For any hiring policy, π, its long-run average service rate is

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

the long-run average number of calls that an agent handles per minute each day. Hence, the quantity $\mu(\pi)^{-1}$ can be used to obtain a rough estimate of the number of agents needed for a given call volume.

The quantities reported in Table 2 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 reported for this example varies slightly (within one standard error) from that reported in Section 7.1.

Table 2   **Comparison with other hiring policies** *(s.e.)*.

Policy	Total Expected Discounted Cost		Fraction of terminated workers	Long-run average service rate	
Optimal policy	5,493.6 *(11.9)*	-	0.3949 *(.0005)*	0.6410 *(.0132)*	-
Never screen	6,050.7 *(15.0)*	10.14%	0.0000 *(.0000)*	0.5361 *(.0141)*	-16.37%
Screen period 1	6,019.7 *(14.5)*	9.58%	0.0195 *(.0002)*	0.5388 *(.0142)*	-15.95%
Screen periods 1 – 5	5,705.6 *(12.9)*	3.86%	0.2076 *(.0005)*	0.5838 *(.0143)*	-8.92%
Screen periods 1 – 10	5,551.1 *(12.4)*	1.05%	0.2991 *(.0005)*	0.6107 *(.0140)*	-4.73%
Screen periods 1 – 20	5,494.2 *(11.8)*	0.01%	0.3556 *(.0005)*	0.6284 *(.0143)*	-1.96%
Gittins policy every 5	5,511.2 *(12.1)*	0.32%	0.3694 *(.0006)*	0.6342 *(.0135)*	-1.06%
Gittins policy every 10	5,575.1 *(11.4)*	1.48%	0.3288 *(.0005)*	0.6218 *(.0139)*	-2.99%
Gittins policy every 20	5,650.4 *(12.2)*	2.85%	0.2617 *(.0005)*	0.6017 *(.0134)*	-6.12%
One-shot decision at 1	5,907.2 *(13.9)*	7.53%	0.2431 *(.0005)*	0.5734 *(.0145)*	-10.54%
One-shot decision at 5	5,619.5 *(12.5)*	2.29%	0.3227 *(.0005)*	0.6114 *(.0149)*	-4.61%
One-shot decision at 10	5,629.3 *(12.3)*	2.47%	0.3234 *(.0006)*	0.6156 *(.0139)*	-3.96%
One-shot decision at 20	5,690.2 *(12.4)*	3.58%	0.2662 *(.0005)*	0.6008 *(.0137)*	-6.27%

The results in Table 2 show that the optimal policy we examined leads to a substantial reduction in cost and to an overall improvement of employee performance. For instance, the policy that does not screen employees has a total expected discounted cost that is 10.14% higher than that of the optimal Gittins-index

policy. Long-run average service rates imply that the optimal Gittins-index policy also leads to a 16.37% decrease in the average number of workers required to maintain the same level of capacity. To more clearly understand this, consider the hypothetical scenario in which the call center has an average load of 53.61 calls per minute. With the optimal policy, this requires to employ 53.61 / 0.6410 = 83.63 workers – long-run average – to have a "fully-loaded" system. With the policy "never screen", the same "fully-loaded" system requires 53.61 / 0.5361 = 100 workers, and the optimal policy employs 16.37% fewer workers.

Recall 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 5 days also performs close to optimally. Thus, screening need not to occur every period in order for a policy to be effective.

"One-shot" hiring and retention decisions are analogues of "secretary"-type rules that limit the frequency of screening as well as the horizon over which screening takes place. We note that the "one-shot decision" policy for period 1 outperforms analogous "screen period 1," because the former uses an optimized stopping boundary, while the latter uses the (unoptimized) stopping boundary from Gittins-index policy in which screening occurs each period. More importantly, the results for "one-shot" 5, 10 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.

7.3. Sensitivity analysis

This section examines how the optimal policy depends on four key parameters: employees' learning rates; employer uncertainty regarding employee performance; task-by-task variability; and hiring costs. The Gittins indices and turnover rates reported in this section are computed as in Section 7.1.

Learning rates. Section 7.1 studied 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 kept constant, as in Section 7.1.

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.

Figure 2 Stopping boundaries for different learning rates.
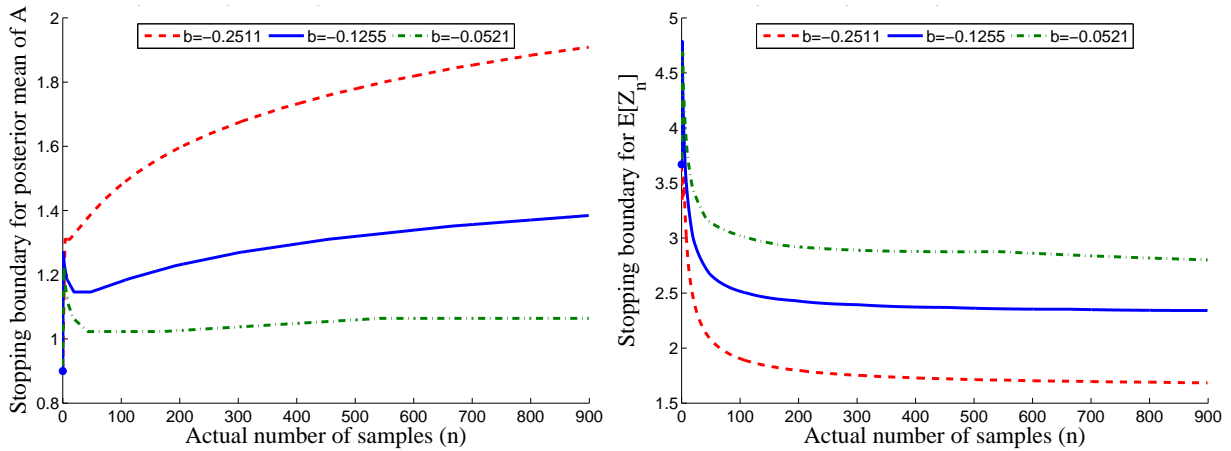
Note. Other parameters: $\mu_0 = 0.90, \sigma_0 = 0.40, \sigma = 0.80, s_0 = 4, c_h = 30, c_f = c_q = 0$.

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 can also look at the 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 the lowest long-run average service rates.

Table 3 Gittins indices, termination and long-run average service rates with different learning rates *(s.e.)*.

| 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 *(.0004)*	0.1834 *(.0017)*	0.0275 *(.0007)*	0.2366 *(.0019)*	1.0253 *(.0286)*
-0.1255	5,491.7	0.0196 *(.0006)*	0.2830 *(.0020)*	0.0557 *(.0010)*	0.3982 *(.0022)*	0.6412 *(.0133)*
-0.0521	6,762.1	0.0334 *(.0008)*	0.3167 *(.0021)*	0.0822 *(.0012)*	0.4885 *(.0022)*	0.4972 *(.0089)*

Table 3's results suggest a potentially important, positive sequence of managerial implications. Improvements in on-the-job learning rates make employees with relatively poor initial abilities quickly become attractive relative to 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 forced turnover 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 in to the realm of stochastic games, however.

Variance of base-level performance in prior distribution. We parameterize the employer's uncertainty concerning the ability of untested workers using the prior variance of A, σ_0^2. By varying σ_0^2, while holding σ^2 constant, we can see how the optimal screening policy changes with worker heterogeneity. Here, we analyze three values of the prior variance, 0.04, 0.36 and 0.64 (i.e., $\sigma_0 = 0.20, 0.40, 0.80$ respectively), and we discuss how they affect our results. All other parameters remain constant, as in Section 7.1.

Table 4   **Gittins indices, termination and long-run average service rates with different prior variances** *(s.e.)*.

| σ_0 | Gittins index | Fraction of terminated workers | | | | Long-run average service rate |
		Day 1	Days 2–10	Days 11 –20	Total	
0.2000	5,715.1	0.0000 *(.0000)*	0.0330 *(.0008)*	0.0351 *(.0008)*	0.1060 *(.0014)*	0.5204 *(.0064)*
0.4000	5,491.7	0.0196 *(.0006)*	0.2830 *(.0020)*	0.0557 *(.0010)*	0.3982 *(.0022)*	0.6412 *(.0133)*
0.8000	4,752.8	0.2406 *(.0019)*	0.2840 *(.0020)*	0.0439 *(.0009)*	0.5842 *(.0022)*	1.1357 *(.0488)*

Table 4 shows how the Gittins index, the fraction of terminated workers, and the long-run average service rate change with σ_0^2. The values obtained in the numerical example agree with the general idea that the Gittins index reflects an option value inherent in the ability to change arms, and it favors arms with more diffuse prior distributions. In our context, this implies that, given constant μ_0, an increase in quality variation across workers allows the employer to screen more strictly, increasing termination rates, retaining relatively more capable employees, and lowering total costs.

Sampling variance. We then perform a sensitivity analysis with respect to the sampling variance σ^2. The analysis is similar to that for the prior variance, but here we keep σ_0 constant as we let σ vary. The values of σ we consider are 0.60, 0.80, 1.00. The other parameters are fixed as in Section 7.1.

Table 5   **Gittins indices, termination and long-run average service rates with different sampling variances** *(s.e.)*.

| σ | Gittins index | Fraction of terminated workers | | | | Long-run average service rate |
		Day 1	Days 2–10	Days 11 –20	Total	
0.6000	4,993.3	0.0057 *(.0003)*	0.1831 *(.0017)*	0.0595 *(.0011)*	0.2755 *(.0020)*	0.6626 *(.0111)*
0.8000	5,491.7	0.0196 *(.0006)*	0.2830 *(.0020)*	0.0557 *(.0010)*	0.3982 *(.0022)*	0.6412 *(.0133)*
1.0000	6,190.4	0.0534 *(.0010)*	0.3426 *(.0021)*	0.0517 *(.0010)*	0.4961 *(.0022)*	0.6104 *(.0160)*

Table 5 displays the increase in the Gittins index and in the long-run average service rate, as σ increases. It also indicates that, for lower σ, the fractions of employees who are terminated are lower. Thus, reductions in within-period variability improve the selectivity and effectiveness of screening procedures, allowing the employer to reduce optimal termination rates.

Hiring costs. Section 7.1 studied a setting in which every time a new worker is employed, the employer incurs a hiring cost, $c_h = 30$. Here we perform a sensitivity analysis that studies how termination rates and total expected discounted costs vary with hiring costs.

Figure 3 plots the stopping boundary with respect to the posterior mean of A and with respect to $\mathbb{E}[Z_n]$ for $c_h = \{0, 15, 30, 60\}$. The left panel shows that the stopping boundary jumps up as hiring costs increase. After day 1, the stopping boundary decreases very slightly for some time, and then it climbs again. The same observations about two competing forces made in Section 7.1 hold here as well.

Figure 3   Stopping boundaries for different hiring costs.
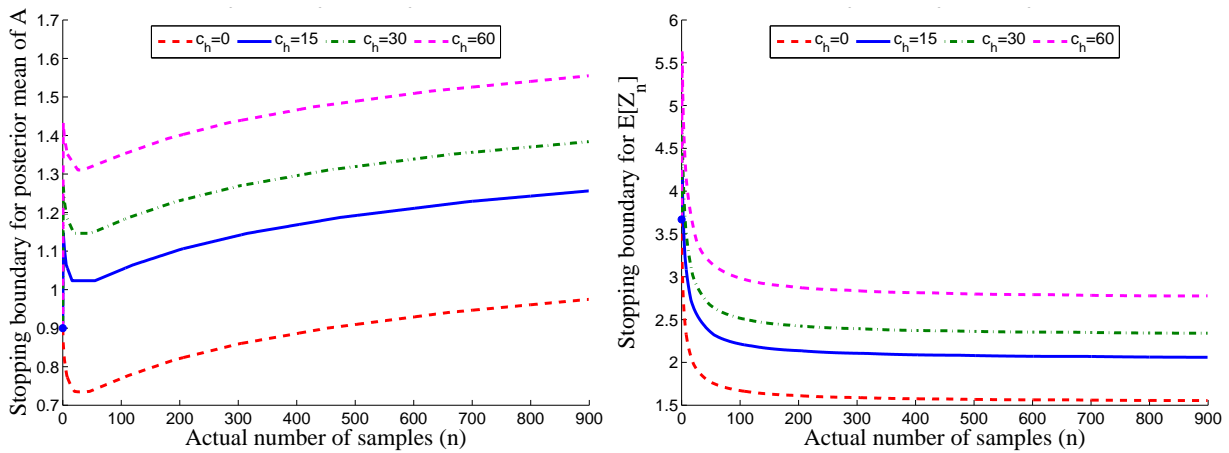
Note. Other parameters: $b = -0.1255, \mu_0 = 0.90, \sigma_0 = 0.40, \sigma = 0.80, s_0 = 4, c_f = c_q = 0$.

When $c_h = 0$, the stopping boundary with respect to $\mathbb{E}[Z_n]$ is monotonically decreasing. This monotonicity is retained for $c_h > 0$, but only after the initial jump attributed to the presence of hiring costs. When hiring costs are absent the screening process is very selective and terminates 58.49% of employees on day 1 and 87.36% overall. As hiring costs enter into the problem, the termination rates quickly drop, and the values of the Gittins indices and of the service rates follow, naturally, the opposite trend.

Table 6   Gittins indices, termination and long-run average service rates with different hiring costs *(s.e.)*.

c_h	Gittins index	Fraction of terminated workers				Long-run average service rate
		Day 1	Days 2–10	Days 11–20	Total	
0	3,645.2	0.5849 *(.0022)*	0.2600 *(.0020)*	0.0153 *(.0005)*	0.8736 *(.0015)*	0.8316 *(.0153)*
15	4,833.7	0.0833 *(.0012)*	0.3844 *(.0022)*	0.0548 *(.0010)*	0.5546 *(.0022)*	0.6889 *(.0138)*
30	5,491.7	0.0196 *(.0006)*	0.2830 *(.0020)*	0.0557 *(.0010)*	0.3982 *(.0022)*	0.6412 *(.0133)*
45	6,028.9	0.0057 *(.0003)*	0.1949 *(.0018)*	0.0525 *(.0010)*	0.2880 *(.0020)*	0.6122 *(.0136)*
60	6,509.1	0.0017 *(.0002)*	0.1404 *(.0016)*	0.0415 *(.0009)*	0.2201 *(.0019)*	0.5949 *(.0134)*

Firing and quitting costs. One would expect that changes in firing 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 hiring, quitting and firing costs affect the Gittins index. To that end, we modify our notation to accounts for these difference, letting $M_i(\nu_{i,n_i}, n_i, 0, c_h, c_f, c_q)$ be the Gittins index (13), and $m_0(c_h, c_f, c_q) = M(\nu_0, 0, 0, c_h, c_f, c_q)$.

THEOREM 3. *If the quitting probabilities $q_{i,n} = q$ for all $i \in \mathcal{S}_0$ and all n then $M_i(\nu_{i,n_i}, n_i, 0, c_h, c_f, c_q) <$ $m_0(c_h, c_f, c_q)$ if and only if $M_i(\nu_{i,n_i}, n_i, 0, c_h + c_f, 0, 0) < m_0(c_h + c_f, 0, 0)$.*

Thus, if the hazard rate for quitting is constant for all employees at all times, then changes in firing 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. But because the relative orderings do not change, changes in the firing and quitting costs do not affect the optimal policy, and we therefore do not report a sensitivity analysis with respect to c_f or c_q.

8. 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 model of the hiring and retention problem integrates aspects from all of these streams. With our solution we prove that Banks and Sundaram's (1992) results for infinite-armed bandits carry over to more complex state spaces, and we show that a "no-recall" property (Corollary 2) ensures that worker lifetimes and costs follow the *iid* pattern of a discounted renewal reward process. The *iid* nature of such a sequence allows us to express the optimal infinite-horizon total expected discounted cost as a function of the Gittins index (Theorem 2).

Our numerical results show that active screening of employees can significantly improve expected costs and long-run average employee performance. Furthermore, because most termination takes place early in employees' tenures, relatively simple finite-horizon and one-shot policies have the potential to perform well. Our sensitivity analysis also 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, reductions in within-task variability and improvements in employee learning both provide the additional benefit of lowering termination rates.

Acknowledgments

References

Aksin, Zeynep, Mor Armony, Vijay Mehrotra. 2007. The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management* **16**(6) 665–688.

Aliprantis, Charalambos D., Kim C. Border. 2006. *Infinite dimensional analysis: A Hitchhiker's Guide*. 3rd ed. Springer, Berlin.

Arlotto, Alessandro, Noah Gans, Stephen E. Chick. 2010. Optimal employee retention when inferring unknown learning curves. *Winter Simulation Conference (WSC), Proceedings of the 2010*. 1178 –1188.

Bailey, Charles D. 1989. Forgetting and the learning curve: a laboratory study. *Management Science* **35**(3) 340–352.

Banks, Jeffrey S., Rangarajan K. Sundaram. 1992. Denumerable-armed bandits. *Econometrica* **60**(5) 1071–1096.

Bergemann, Dirk, Juuso Välimäki. 2001. Stationary multi-choice bandit problems. *Journal of Economic Dynamics and Control* **25**(10) 1585 – 1594.

Berry, Donald A., Bert Fristedt. 1985. *Bandit problems: Sequential allocation of experiments*. Monographs on Statistics and Applied Probability, Chapman & Hall, London.

Bertsekas, Dimitri P., Steven E. Shreve. 1996. *Stochastic optimal control: The discrete time case*. Athena Scientific, Belmont.

Billingsley, Patrick. 1968. *Convergence of probability measures*. John Wiley & Sons Inc., New York.

Brown, Lawrence, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, Linda Zhao. 2005. Statistical analysis of a telephone call center: a queueing-science perspective. *J. Amer. Statist. Assoc.* **100**(469) 36–50.

Carillo, Janice E., Cheryl Gaimon. 2000. Improving manufacturing performance through process change and knowledge creation. *Management Science* **46**(2) 265–288.

Dada, Maqbool, K. N. Srikanth. 1990. Monopolistic Pricing and the Learning Curve: An Algorithmic Approach. *Operations Research* **38**(4) 656–666.

Easley, David, Nicholas M. Kiefer. 1988. Controlling a stochastic process with unknown parameters. *Econometrica* **56**(5) 1045–1064.

Farias, Vivek F., Ritesh Madan. 2011. The irrevocable multi-armed bandit problem. *Operations Research* **59**(2) 383–399.

Frazier, Peter, Warren Powell, Savas Dayanik. 2009. The knowledge-gradient policy for correlated normal beliefs. *INFORMS J. Comput.* **21**(4) 599–613.

Freeman, P. R. 1983. The secretary problem and its extensions: a review. *Internat. Statist. Rev.* **51**(2) 189–206.

Gaimon, Cheryl. 1997. Planning information technology-knowledge worker systems. *Management Science* **43**(9) 1308–1328.

Gaimon, Cheryl, Gulru F. Ozkan, Karen Napoleon. to appear. Dynamic resource capabilities: Managing workforce knowledge with a technology upgrade. *Organization Science* .

Gans, Noah, Ger Koole, Avishai Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* **5**(2) 79–141.

Gans, Noah, Nan Liu, Avishai Mandelbaum, Haipeng Shen, Han Ye. 2010. Service times in call centers: Agent heterogeneity and learning with some operational consequences. *Borrowing Strength: Theory Powering Applications – A Festschrift for Lawrence D. Brown*, *Collections*, vol. 6. Inst. Math. Statist., Beachwood, OH, 99–123.

Gans, Noah, Yong-Pin Zhou. 2002. Managing learning and turnover in employee staffing. *Operations Research* **50**(6) 991–1006.

Gittins, John C. 1989. *Multi-Armed Bandit Allocation Indices*. John Wiley & Sons, Chichester, England.

Goldberg, Matthew S., Anduin E. Touw. 2003. *Statistical Methods for Learning Curves and Cost Analysis*. INFORMS. Topics in Operations Research Series.

Huselid, Mark A. 1995. The impact of human resource practices on turnover, productivity, and corporate financial performance. *Academy of Management Journal* **38**(3) 635 – 672.

Jovanovic, Boyan. 1979. Job matching and the theory of turnover. *Journal of Political Economy* **87**(5) 972.

Lapré, Michael A., Amit Shankar Mukherjee, Luk N. Van Wassenhove. 2000. Behind the learning curve: Linking learning activities to waste reduction. *Management Science* **46**(5) 597–611.

Mazzola, Joseph B., Kevin F. McCardle. 1996. A Bayesian approach to managing learning-curve uncertainty. *Management Science* **42**(5) 680–692.

Mazzola, Joseph B., Kevin F. McCardle. 1997. The stochastic learning curve: Optimal production in the presence of learning-curve uncertainty. *Operations Research* **45**(3) 440–450.

Müller, Alfred, Dietrich Stoyan. 2002. *Comparison methods for stochastic models and risks*. Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester.

Nagypál, Éva. 2007. Learning by doing vs. learning about match quality: Can we tell them apart? *Review of Economic Studies* **74** 537–566.

Nembhard, David A. 2001. Heuristic approach for assigning workers to tasks based on individual learning rates. *International Journal of Production Research* **39** 1955–1968(14).

Nembhard, David A., Napassavong Osothsilp. 2002. Task complexity effects on between-individual learning/forgetting variability. *International Journal of Industrial Ergonomics* **29** 297–306(10).

Nembhard, David A., Mustafa V. Uzumeri. 2000a. Experiential learning and forgetting for manual and cognitive tasks. *International Journal of Industrial Ergonomics* **25** 315–326(12).

Nembhard, David A., Mustafa V. Uzumeri. 2000b. An individual-based description of learning within an organization. *IEEE Transactions on Engineering Management* **47**(3) 370–378.

Pinker, Edieal J., Robert A. Shumsky. 2000. The efficiency-quality trade-off of cross-trained workers. *Manufacturing & Service Operations Management* **2**(1) 32–48.

Pisano, Gary P., Richard M.J. Bohmer, Amy C. Edmondson. 2001. Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science* **47**(6) 752–768.

Rudin, Walter. 1976. *Principles of mathematical analysis*. 3rd ed. McGraw-Hill Book Co., New York. International Series in Pure and Applied Mathematics.

Shafer, Scott M., David A. Nembhard, Mustafa V. Uzumeri. 2001. The effects of worker learning, forgetting, and heterogeneity on assembly line productivity. *Management Science* **47**(12) 1639–1653.

Shaked, Moshe, J. George Shanthikumar. 2007. *Stochastic orders*. Springer Series in Statistics, Springer, New York.

Shen, Haipeng. 2003. *Estimation, Confidence Intervals and Nonparametric Regression for Problems Involving Log-normal Distributions*. Ph.D. Thesis, University of Pennsylvania.

Shen, Haipeng, Lawrence D. Brown. 2006. Non-parametric modelling for time-varying customer service time at a bank call centre. *Appl. Stoch. Models Bus. Ind.* **22**(3) 297–311.

Whitt, Ward. 2006. The impact of increased employee retention on performance in a customer contact center. *Manufacturing Service Operations 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, Eelke. 2007. Conditions that shape the learning curve: Factors that increase the ability and opportunity to learn. *Management Science* **53**(12) 1903–1915.

Yelle, Louis E. 1979. The learning curve: historical review and comprehnsive survey. *Decision Sciences* **10**(2) 302–328.

Appendix. Mathematical Results

Proofs of mathematical claims are presented in the order of their appearance in the main paper. (The statement "Proof of …" is presented in bold face). When other technical results are needed, they are stated with a full proof or suitable reference, in the location that they are needed (the result is presented in standard typeface).

Proof of Lemma 1. In this proof of (i) we assume $c_h = c_f = c_q = 0$, and we leave to the interested reader to check the details for the general case. Suppose there is a policy π_2 that uses at least one worker i at least once when his experience n_i exceeds Λ_i, with positive probability. Define the random variable $\bar{T} = \inf\{t : n_{\pi_2(t)} \geq \Lambda_{\pi_2(t)}\}$ to be the first time that such a worker is employed (with $\bar{T} = \infty$ on sample paths where an employee is not use in that way). By assumption, then, $\mathbb{P}\left(\bar{T} < \infty\right) > 0$. The cost at time $\bar{T}$ is $\gamma^{\bar{T}} c(K)$.

We define a new policy, π_1, that is nonanticipative if π_2 is, and that has a lower expected total discounted cost. Let $\pi_1(t) = \pi_2(t)$ for $t < \bar{T}$, and let $\pi_1(t) = \pi_2(t+1)$ for $t \geq \bar{T}$. Then

$$C_{\pi_1} = \sum_{t=0}^{\bar{T}-1} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right) + \sum_{t=\bar{T}}^{\infty} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right)$$

$$C_{\pi_2} = \sum_{t=0}^{\bar{T}-1} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right) + \gamma^{\bar{T}} c(K) + \sum_{t=\bar{T}+1}^{\infty} \gamma^t c\left(Z(\nu_{\pi_1(t-1)}, n_{\pi_1(t-1)})\right).$$

We prove the statement showing that $C_{\pi_1} - C_{\pi_2} < 0$ when $\bar{T} < \infty$.

$$\begin{aligned}
C_{\pi_1} - C_{\pi_2} &= \sum_{t=0}^{\bar{T}-1} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right) + \sum_{t=\bar{T}}^{\infty} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right) \\
&\quad - \sum_{t=0}^{\bar{T}-1} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right) - \gamma^{\bar{T}} c(K) - \sum_{t=\bar{T}+1}^{\infty} \gamma^t c\left(Z(\nu_{\pi_1(t-1)}, n_{\pi_1(t-1)})\right) \\
&= (1-\gamma)\left(\sum_{t=\bar{T}}^{\infty} \gamma^t c\left(Z(\nu_{\pi_1(t)}, n_{\pi_1(t)})\right)\right) - \gamma^{\bar{T}} c(K) \\
&< (1-\gamma)\left(\sum_{t=\bar{T}}^{\infty} \gamma^t c(K)\right) - \gamma^{\bar{T}} c(K) = (1-\gamma)\frac{\gamma^{\bar{T}} c(K)}{1-\gamma} - \gamma^{\bar{T}} c(K) = 0
\end{aligned}$$

Thus, π_2 is strictly worse than π_1 on a set of nonzero probability, and is therefore not optimal, as was to be shown.

We now prove part (ii) of the lemma.

If: Let π^* be optimal for Problem 2. Then by part (i) of the lemma, there is no worker who is retained so that $n_i \geq \Lambda_i$. Therefore, π^* is feasible for Problem 1, too. For any π, let C^1_π, C^2_π be the total expected cost of policy π in Problem 1 and 2 and observe that $C^1_\pi = C^2_\pi$ for any π. Then, for any $\pi \neq \pi^*$, $C^1_\pi = C^2_\pi \geq C^2_{\pi^*} = C^1_{\pi^*}$. Therefore, π^* is also optimal for Problem 1.

Only if: Let π^* be optimal for Problem 1. Then, any policy that is feasible for Problem 1 is feasible for Problem 2. Observing again that $C^1_\pi = C^2_\pi$ for any π and that for any $\pi \neq \pi^*$, $C^2_\pi = C^1_\pi \geq C^1_{\pi^*} = C^2_{\pi^*}$ it follows that π^* is optimal also for Problem 2.

Proof of Lemma 2. Suppose that $\pi_1 \in \Pi$ 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 $\Lambda_i(\pi_1)$ in (3) satisfies $0 \leq \Lambda_i(\pi_1) \leq \Lambda_i$ on every sample path, for all $i \in \mathcal{S}_0$. Suppose, by contradiction, that policy $\pi_1 \in \Pi$ only employs $\kappa_1 < \infty$ workers with some positive probability $\epsilon > 0$. Then, since $\pi_1 \in \Pi$ we have

$$\mathbb{P}\left(\sum_{i=1}^{\kappa_1} \Lambda_i(\pi_1) \geq \zeta\right) \geq \epsilon \tag{17}$$

for all $\zeta \in \mathbb{R}$. Since $0 \leq \Lambda_i(\pi_1) \leq \Lambda_i$, together with Markov's inequality, we obtain $\mathbb{P}\left(\sum_{i=1}^{\kappa_1} \Lambda_i(\pi_1) \geq \zeta\right) \leq \mathbb{P}\left(\sum_{i=1}^{\kappa_1} \Lambda_i \geq \zeta\right) \leq \kappa_1 \mathbb{E}[\Lambda_1]/\zeta$. Selecting any $\zeta > \kappa_1 \mathbb{E}[\Lambda_1]/\epsilon$, we obtain a contradiction to (17) from which we can conclude that $\pi_1 \notin \Pi$. Hence, each policy for Problem 1 employs an infinite number of workers with probability 1.

Proof of Theorem 1. We turn now to the proof of Theorem 1. The argument requires a delicate analysis of the Bellman equation (11) and of the Gittins index (13) whose properties are presented in order. In our analysis, we use the weak topology on $\mathcal{P}(\Omega)$. Thus, a sequence of measures $\{\nu_t\}_{t=1}^{\infty}$ in $\mathcal{P}(\Omega)$ converges to a measure ν if and only if $\int_\Omega h(\boldsymbol{\theta}) d\nu_t \to \int h(\boldsymbol{\theta}) d\nu$ for every continuous, real-valued function h on Ω. In addition, we equip the space $\mathcal{P}(\Omega)$ with the the Prokhorov distance function, ϱ, so that the space $(\mathbb{P}(\Omega), \varrho)$ is a metric space. (see, e.g. Billingsley 1968 and Aliprantis and Border 2006, Chapter 15).

LEMMA 3. (i) *For each ν, n and r, $V(\nu, m, n, r)$ is concave, non-decreasing and Lipschitz continuous in m, with Lipschitz constant equal to 1.* (ii) *For each m, n, r, $V(\nu, m, n, r)$ is continuous in ν.*

Proof. (i) When $r = 1$, the result is trivial and its proof is omitted. When $r = 0$, we proceed by means of the Value Iteration Algorithm (see, e.g. Bertsekas and Shreve 1996, Section 9.5, Definition 9.10 and Proposition 9.14).

Let $v^0(\nu, m, n, 0) = 0$ for all $m \in \mathbb{R}$, and notice that v^0 is trivially nondecreasing, concave, and Lipschitz-1 continuous in m for each ν, n. Assume that $v^{k-1}(\nu, m, n, 0)$ is nondecreasing, concave, and Lipschitz-1 continuous in m for each ν, n. Let

$$v^k(\nu, m, n, 0) = \min\{c_f + m, \tag{18}$$
$$c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))] + \gamma(1 - q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), m, n + 1, 0)] + \gamma q_n V(\mathbb{1}_K, m, n + 1, 1)\},$$

and notice that $c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))]$ is constant with respect to m, $\gamma q_n V(\mathbb{1}_K, m, n + 1, 1)$ is nondecreasing, concave, and Lipschitz-γq_n continuous in m, and $\gamma(1 - q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), m, n + 1, 0)]$ is nondecreasing, concave, and Lipschitz-$\gamma(1 - q_n)$ continuous in m by the induction assumption. Recalling that monotonicity and concavity are preserved under minimization, we obtain that $v^k(\nu, m, n, 0)$ is nondecreasing and concave in m.

To obtain that $v^k(\nu, m, n, 0)$ 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. Since $\gamma < 1$, the result follows.

(ii) When $r = 1$, the Bellman equation does not depend on ν and the result is then trivial. When $r = 0$, we proceed by means of the Value Iteration Algorithm. Let $v^0(\nu, m, n, 0) = 0$ for all $\nu \in \mathcal{P}(\Omega)$ and notice that v^0 is continuous in ν, for any given m, n. Assume now that $v^{k-1}(\nu, m, n, 0)$ is continuous in ν and let

$$v^k(\nu, m, n, 0) = \min\{c_f + m,$$
$$c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))] + \gamma(1 - q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), m, n + 1, 0)] + \gamma q_n V(\mathbb{1}_K, m, n + 1, 1)\},$$

For $\nu_t \to \nu$ we have that

$$\mathbb{E}[c(Z(\nu_t, n))] = \int_\Omega \left(\int_{K_{\text{inf}}}^{K^{\text{sup}}} c(z)\xi_n(z|\boldsymbol{\theta})dz \right) d\nu_t \to \int_\Omega \left(\int_{K_{\text{inf}}}^{K^{\text{sup}}} c(z)\xi_n(z|\boldsymbol{\theta})dz \right) d\nu = \mathbb{E}[Z(\nu, n)],$$

by the definition of the weak topology on $\mathcal{P}(\Omega)$ since the $f(\boldsymbol{\theta}) = \int_{K_{\text{inf}}}^{K^{\text{sup}}} c(z)\xi_n(z|\boldsymbol{\theta})dz$ is continuous in $\boldsymbol{\theta}$ by the joint continuity of $\xi_n(z|\boldsymbol{\theta})$ in $(z, \boldsymbol{\theta})$. Interchanges of limits and integrals are justified by the Dominated Convergence Theorem. In addition, Easley and Kiefer (1988, Theorem 1) also show that, under our assumptions on $\xi_n(z|\boldsymbol{\theta})$, the Bayes operator is continuous in ν. Hence, for $\nu_t \to \nu$, $\beta(\nu_t, z) \to \beta(\nu, z)$. Since v^{k-1} is continuous in ν by the induction assumption and the composition of two continuous functions is again continuous we then obtain that $v^{k-1}(\beta(\nu, z), m, n + 1, 0)$ is continuous in ν for any given z. Then, the weak topology on $\mathcal{P}(\Omega)$ and the Dominated Convergence Theorem yield that $\mathbb{E}[v^{k-1}(\beta(\nu_t, Z(\nu_t, n)), m, n + 1, 0)] \to \mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), m, n + 1, 0)]$ proving continuity of the second addend. The third addend and the first minimand are constant, hence continuous, with respect to ν. The result then follows since continuity is closed with respect to addition and minimization. $\square$

LEMMA 4. *For any given n, r, $V(\nu, m, n, r)$ is jointly continuous in ν, and m.*

Proof. For $r = 1$, the statement is equivalent to the continuity in m that was proven in Lemma 3. For $r = 0$, observe that $V(\nu, m, n, 0)$ is continuous in ν by Lemma 3–(ii). By Lemma 3–(i) we know that it is also Lipschitz continuous, hence continuous in m. Let $\{(\nu_t, m_t)\}_{t=1}^\infty$ be a sequence in $\mathcal{P}(\Omega) \times \mathbb{R}$ and define a distance function $\hat{\varrho} : (\mathcal{P}(\Omega) \times \mathbb{R})^2 \to \mathbb{R}_+$ as $\hat{\varrho}((\nu, m), (\nu', m')) = \varrho(\nu, \nu') + |m - m'|$, where $\varrho(\cdot, \cdot)$ is the Prokhorov metric on $\mathcal{P}(\Omega)$ induced by the weak topology. It is easy to see that $\hat{\varrho}$ is an appropriate distance function on the product space $\mathcal{P}(\Omega) \times \mathbb{R}$. Then, for $(\nu_t, m_t) \to (\nu, m)$,

$$|V(\nu_t, m_t, n, 0) - V(\nu, m, n, 0)| \le |V(\nu_t, m_t, n, 0) - V(\nu_t, m, n, 0)| + |V(\nu_t, m, n, 0) - V(\nu, m, n, 0)| \to 0$$

since, for any given n, the first addend converges to 0 by the continuity in m for any given ν (Lemma 3–(i)) and the second addend converges to 0 by the continuity in ν for any given m (Lemma 3–(ii)). $\square$

LEMMA 5.   (i) $c_f + m > M(\nu, n, 0)$ *if and only if* $HV(\nu, m, n) < c_f + m$.

(ii) $c_f + m < M(\nu, n, 0)$ *if and only if* $HV(\nu, m, n) > c_f + m$.

(iii) $c_f + m = M(\nu, n, 0)$ *if and only if* $HV(\nu, m, n) = c_f + m$.

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

(i) If $HV(\nu, m, n) < c_f + m$, it follows that $V(\nu, m, n, 0) < c_f + m$. Recall that $M(\nu, n, 0) = \sup\{\widetilde{m} \in \mathbb{R} : V(\nu, \widetilde{m}, n, 0) = c_f + \widetilde{m}\}$ and that $V(\nu, \widetilde{m}, n, 0)$ is concave and non-decreasing in $\widetilde{m}$, and $V(\nu, \widetilde{m}, n, 0) \le c_f + \widetilde{m}$ for all $\widetilde{m}$. This implies $c_f + m > M(\nu, n, 0)$.

Again, recall that $M(\nu, n, 0) = \sup\{\widetilde{m} \in \mathbb{R} : V(\nu, \widetilde{m}, n, 0) = c_f + \widetilde{m}\}$ and that $V(\nu, \widetilde{m}, n, 0)$ is concave and non-decreasing in $\widetilde{m}$. If $c_f + m > M(\nu, n, 0)$ then $V(\nu, m, n, 0) < c_f + m$. In turn, $V(\nu, m, n, 0) < c_f + m$ implies it is optimal not to retire so $HV(\nu, m, n) = V(\nu, m, n, 0) < c_f + m$.

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

(iii) It follows combining claims (i) and (ii). $\square$

LEMMA 6. *For any n and any r, $M(\nu, n, r)$ is uniformly continuous in ν.*

Proof. For $r = 1$, $M(\nu, n, 1) = \sup\{m \in \mathbb{R} : \min\{c_q + m, c(K) + \gamma V(\mathbb{1}_K, m, n+1, 1)\} = c_q + m\}$ which is constant with respect to ν. Hence, the result is trivial. For $r = 0$, we first show that $M(\nu, n, 0)$ is continuous in ν for any given n. Let $\nu_t \to \nu$ and define $m_t = M(\nu_t, n, 0)$. Since $M(\cdot)$ takes values in a compact set, we can extract a subsequence of m_t (also denoted by m_t) such that $m_t \to m$. By joint continuity (Lemma 4) of $V(\nu, m, n, 0)$ in ν and m, it follows that $V(\nu_t, m_t, n, 0) \to V(\nu, m, n, 0)$. Similarly, $HV(\nu_t, m_t, n) \to HV(\nu, m, n)$. By Lemma 5-(iii), we have $c_f + m_t = HV(\nu_t, m_t, n) = V(\nu_t, m_t, n, 0)$ for any t. Hence, $c_f + m = HV(\nu, m, n) = V(\nu, m, n, 0)$, which, by Lemma 5-(iii), implies $M(\nu, n, 0) = c_f + m$, establishing continuity. To see uniform continuity, recall that for any given n, $M : \mathcal{P}(\Omega) \to [cK_{\inf}/(1-\gamma), cK/(1-\gamma)]$. Since Ω is compact, it follows that $\mathcal{P}(\Omega)$ is compact as well (see, e.g. Aliprantis and Border 2006, Theorem 15.11). The space $\mathcal{P}(\Omega)$ equipped with the Prokhorov distance is then a compact metric space, and $M(\cdot)$ is a continuous map of a compact metric space that is a fortiori uniformly continuous (see, e.g. Rudin 1976, Theorem 4.19). $\square$

We now recall the main result of Banks and Sundaram (1992).

THEOREM 4 **(Banks and Sundaram 1992, Theorem 4.1)**. *Let ν_{i,n_i} have finite support for all i, n_i. Then, worker i with age n_i and state r_i is an optimal selection at time t if and only if $M_i(\nu_{i,n_i}, n_i, r_i) = \inf_j\{M_j(\nu_{j,n_j}, n_j, r_j)\}$, and there exists an optimal policy that always selects a Gittins-index-minimal worker.*

Define $\mathrm{co}(\Omega) = \{\nu \in \mathcal{P}(\Omega) : |\mathrm{supp}(\nu)| < \infty\}$ to be the set of probability measures, ν, defined on the Borel sets of Ω that have finite support. It turns out that $\mathrm{co}(\Omega)$ is dense in $\mathcal{P}(\Omega)$.

THEOREM 5 **(see, e.g. Aliprantis and Border 2006, Theorem 15.10 or Billingsley 1968, Appendix III, Theorem 4)**. *If Ω is metrizable, then $\mathrm{co}(\Omega)$, the set of probability measures with finite support, is dense in $\mathcal{P}(\Omega)$.*

Hence, for every $\nu \in \mathcal{P}(\Omega)$ and $\delta > 0$, there exists $\widetilde{\nu} \in \mathrm{co}(\Omega)$ so that $\varrho(\nu, \widetilde{\nu}) < \delta$, where $\varrho(\cdot, \cdot)$ is the Prokhorov metric on $\mathcal{P}(\Omega)$.

THEOREM 6. *For every $\nu_{i,n_i} \in \mathcal{P}(\Omega)$ there is $\widetilde{\nu}_{i,n_i} \in \mathrm{co}(\Omega)$ such that $M_i(\nu_{i,n_i}, n_i, r_i) = \inf_j\{M_j(\nu_{j,n_j}, n_j, r_j)\}$ if and only if $M_i(\widetilde{\nu}_{i,n_i}, n_i, r_i) = \inf_j\{M_j(\widetilde{\nu}_{j,n_j}, n_j, r_j)\}$.*

Proof of Theorem 6. *If:* Fix $\epsilon > 0$. Then, since $M(\nu, n, r)$ is uniformly continuous in ν for any given n and r (see Lemma 6), there exists $\delta > 0$ such that

$$|M(\nu, n, r) - M(\nu', n, r)| < \frac{\epsilon}{2}. \tag{19}$$

for any pair ν, ν' such that $\varrho(\nu, \nu') < \delta$. By Theorem 5 we can choose $\nu \equiv \nu_{i,n_i} \in \mathcal{P}(\Omega)$ to be the type distribution of worker i with experience n_i, and $\nu' \equiv \widetilde{\nu}_{i,n_i} \in \mathrm{co}(\Omega)$ to be the finitely supported equivalent of ν_{i,n_i}. Hence, (19) implies that

$$|M(\nu_{i,n_i}, n_i, r_i) - M(\widetilde{\nu}_{i,n_i}, n_i, r_i)| < \frac{\epsilon}{2}. \tag{20}$$

Now, let $M_i(\widetilde{\nu}_{i,n_i}) = \inf_j\{M_j(\widetilde{\nu}_{j,n_j})\}$, and notice that $M_i(\nu_{i,n_i}, n_i, r_i) \geq \inf_j\{M_j(\nu_{j,n_j}, n_i, r_i)\}$. Then, suppressing the n_i's and the r_i's, we have

$$M_i(\nu_{i,n_i}) \leq |M_i(\nu_{i,n_i}) - M_i(\widetilde{\nu}_{i,n_i})| + M_i(\widetilde{\nu}_{i,n_i}) < \frac{\epsilon}{2} + \inf_j\{M_j(\widetilde{\nu}_{j,n_j})\} \tag{21}$$

$$\leq \frac{\epsilon}{2} + \inf_j\{|M_j(\widetilde{\nu}_{j,n_j}) - M_j(\nu_{j,n_j})| + M_j(\nu_{j,n_j})\} < \epsilon + \inf_j\{M_j(\nu_{j,n_j}))\}, \tag{22}$$

where (21) and (22) follow from (19). Letting $\epsilon \to 0$ yields $M_i(\nu_{i,n_i}) = \inf_j\{M_j(\nu_{j,t})\}$, as required for the "if" part of the proof. A detailed "only if" part of the proof is omitted for brevity – it is very similar to the above "if" proof (the roles of ν and $\widetilde{\nu}$ are reversed.) $\square$

The proof of Theorem 1 follows now easily.

Proof of Theorem 1. Combine Theorem 6, which shows that the probability distributions with finite support are dense in the set of probability measures on Ω, with Banks and Sundaram (1992, Theorem 4.1), which shows the asserted result holds true for the case of measures with finite support. $\square$

Proof of Corollary 1. At $t = 0$, no worker has ever been employed and all the workers have Gittins index m_0. Then, the sampling process starts with a random selection of worker, i, from the stationary pool of candidates. Worker i is employed as long as $M_i(\nu_{i,n_i}, n_i, r_i) = \inf_j\{M_j(\nu_{j,t}, n_j, r_j)\} \leq m_0$. As soon as i is discarded, $M_i(\nu_{i,n_i}, n_i, r_i) > m_0$ and the sampling process starts again.

Proof of Corollary 2. It follows immediately from Lemma 1 and Theorem 1.

Proof of Theorem 2. Consider the retirement-option problem described in Section 4. By Lemma 5-(iii), we obtain

$$c_f + m_0 = HV(\nu_0, m_0, 0), \tag{23}$$

and we note that $HV(\nu_0, m_0, 0)$ is the total expected discounted cost of the retirement-option problem. Further, we let $\widetilde{\Lambda} = \inf\{n \in \mathbb{N} : HV(\nu_n, m_0, n) = c_f + m_0\}$ be the optimal stopping time for which the employer decides to end her business and pay the terminal cost m_0, and we note that $\widetilde{\Lambda} \stackrel{d}{=} 1 + \Lambda_i(\pi^*)$. Then, $HV(\nu_0, m_0, 0, 0) = c_h + \mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^t c(Z(\nu_t, t)) + \gamma^{\widetilde{\Lambda}}(c_f \mathbb{1}(\widetilde{\Lambda} < \Lambda) + c_q \mathbb{1}(\widetilde{\Lambda} = \Lambda) + m_0)\right]$, and by using (23), we obtain

$$\left(1 - \mathbb{E}\left[\gamma^{\widetilde{\Lambda}}\right]\right)(c_f + m_0) = c_h + \mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^t c(Z(\nu_t, t)) + \gamma^{\widetilde{\Lambda}}(c_q - c_f)\mathbb{1}(\widetilde{\Lambda} = \Lambda)\right]. \tag{24}$$

By Theorem 1, we know that the optimal policy for our original infinite-horizon-expected-discounted-cost problem requires that the employer retains the jth worker as long as he is Gittins-index minimal, i.e. $M_j(\nu_n, n, r_n) \leq m_0$. Let $\widetilde{\Lambda}_j = \inf\{n \in \mathbb{N} : M_j(\nu_n, n, 0) > m_0\}$ represent the tenure, under the optimal policy π^*, at which the jth worker becomes unavailable (either because he resigned or because he was terminated). Set $\widetilde{\Lambda}_0 = 0$, and notice that $\{\widetilde{\Lambda}_j\}_{j=1}^{\infty}$ is the *iid* sequence of optimal working-life times of the employees. Because $\mathbb{E}[\Lambda_j] < \infty$ and the Λ_j are *iid*, Lemma 2 implies that the optimal policy will employ an a.s. infinite number of workers. Moreover,

$$
\begin{aligned}
\inf_{\pi \in \Pi} C_\pi(\nu_0) &= \mathbb{E}\left[\sum_{k=1}^{\infty} \gamma^{\sum_{j=0}^{k-1} \widetilde{\Lambda}_j} \sum_{t=0}^{\widetilde{\Lambda}_k - 1} \gamma^t \left\{c_h \mathbb{1}(t=0) + c(Z(\nu_{\pi(t)}, n_{\pi(t)}))\right\} + \gamma^{\widetilde{\Lambda}_k}(c_q - c_f)\mathbb{1}(\widetilde{\Lambda}_k = \Lambda_k) + \gamma^{\widetilde{\Lambda}_k} c_f\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}\left[\gamma^{\sum_{j=0}^{k-1} \widetilde{\Lambda}_j}\right] \mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}_k - 1} \gamma^t \left\{c_h \mathbb{1}(t=0) + c(Z(\nu_{\pi(t)}, n_{\pi(t)}))\right\} + \gamma^{\widetilde{\Lambda}_k}(c_q - c_f)\mathbb{1}(\widetilde{\Lambda}_k = \Lambda_k) + \gamma^{\widetilde{\Lambda}_k} c_f\right] \\
&= (c_f + m_0)\left(1 - \mathbb{E}\left[\gamma^{\widetilde{\Lambda}_1}\right]\right) \sum_{k=1}^{\infty} \mathbb{E}\left[\gamma^{\sum_{j=0}^{k-1} \widetilde{\Lambda}_j}\right] + c_f \sum_{k=1}^{\infty} \mathbb{E}\left[\gamma^{\sum_{j=0}^{k} \widetilde{\Lambda}_j}\right] \\
&= m_0 + c_f\left(1 - \mathbb{E}\left[\gamma^{\widetilde{\Lambda}_1}\right]\right)^{-1} = m_0 + c_f\left(1 - \gamma \mathbb{E}\left[\gamma^{\Lambda_1(\pi^*)}\right]\right)^{-1},
\end{aligned}
\tag{25}
$$

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

Proof of Proposition 1. To prove Proposition 1, we first prove the following lemma. In it, we use the notion of a likelihood ratio order (see Shaked and Shanthikumar 2007, Section 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 $X \leq_{\mathrm{lr}} Y$ (X is stochastically smaller than Y in the likelihood ratio sense) if $f_Y(t)/f_X(t)$ increases in t over the union of the supports of X and Y.

LEMMA 7. *Let $g : \mathbb{R}^3 \to \mathbb{R}$ be such that (i) for $A \sim \nu$, $\beta(\nu, z)([-\infty, a]) = \mathbb{P}(A \leq a|Z = z)$ is nondecreasing in z for any given ν. If, for any $a \leq a'$, $\xi_n(z|a')/\xi_n(z|a)$ is nondecreasing in z, then $V(\nu, m, n, r) \leq V(\nu', m, n, r)$, for any $\nu \leq_{\mathrm{lr}} \nu'$, and for each given m, n, r.*

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

Proof of Lemma 7. For $r = 1$, $V(\nu, m, n, 1) = \min\{c_q + m, c(K) + \gamma V(\mathbb{1}_K, m, n+1, 1)\}$, which is constant and hence trivially nondecreasing in ν. For $r = 0$, he assumption that $\xi_n(z|a')/\xi_n(z|a)$ is nondecreasing in z for any $a \leq a'$ yields that $Z(a, n) \leq_{\mathrm{lr}} Z(a', n)$. Then, since $\nu \leq_{\mathrm{lr}} \nu'$, Shaked and Shanthikumar (2007, Theorem 1.C.17) yield that $Z(\nu, n) \leq_{\mathrm{lr}} Z(\nu', n)$. We can now show monotonicity of the Bellman equation (11) with respect to the likelihood ratio order. We proceed by means of the Value Iteration Algorithm (see, e.g. Bertsekas and Shreve 1996, Section 9.5, Definition 9.10 and Proposition 9.14). Let $v^0(\nu, m, n, 0) = 0$ for all distributions ν, and notice that v^0 is trivially nondecreasing in ν. Assume $v^{k-1}(\nu, m, n, 0) \leq v^{k-1}(\nu', m, n, 0)$ for $\nu \leq_{\mathrm{lr}} \nu'$. Then,

$$v^k(\nu, m, n, 0) = \min\{c_f + m, \tag{26}$$

$$c_h \mathbb{1}(n = 0) + \mathbb{E}[c(Z(\nu, n))] + \gamma(1 - q_n)\mathbb{E}[v^{k-1}(\beta(\nu, Z(\nu, n)), m, n+1, 0)] + \gamma q_n V(\mathbb{1}_K, m, n+1, 1)\},$$

Notice that $\mathbb{E}[v^{k-1}(\beta(\nu_t, Z(\nu_t, n)), m, n+1, 0)]$ is an expectation with respect to Z, and recall that $Z(\nu, n) \leq_{\mathrm{lr}} Z(\nu', n)$, and that, by Condition (i), $\beta(\nu, z)$ is nondecreasing in z. Recall that $\leq_{\mathrm{lr}} \Rightarrow \leq_{\mathrm{st}}$ (Shaked and Shanthikumar 2007, Theorem 1.C.1), where $\leq_{\mathrm{st}}$ is the usual stochastic order. Then, Shaked and Shanthikumar (2007, Example 1.C.57) and the induction assumption yield $\mathbb{E}[v^{k-1}(\beta(\nu_t, Z(\nu_t, n)), m, n+1, 0)] \leq \mathbb{E}[v^{k-1}(\beta(\nu'_t, Z(\nu'_t, n)), m, n+1, 0)]$. Also, $\mathbb{E}[Z(\nu_t, n)] \leq \mathbb{E}[Z(\nu'_t, n)]$, showing that the second minimand in (26) is nondecreasing in ν.

The first minimand in (26) is constant for any given m, and monotonicity is preserved under minimization. Hence, $v^k(\nu, m, n, 0) \leq v^k(\nu', m, n, 0)$, for any $\nu \leq_{\mathrm{lr}} \nu'$. Repeated application of the Value Iteration Algorithm then yields $V(\nu, m, n, r) \leq V(\nu', m, n, r)$, for any $\nu \leq_{\mathrm{lr}} \nu'$, as desired. $\square$

Proof of Proposition 1. The posterior distribution of A has distribution $N(w_s, \sigma^2/s)$. The normal distribution has the monotone likelihood ratio property required by Lemma 7 (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. When $r = 1$, $M(\nu, n, 1) = \sup\{m \in \mathbb{R} : \min\{c_q + m, c(K) + \gamma V(\mathbb{1}_K, m, n+1, 1)\} = c_q + m\}$, which is constant and hence trivially nondecreasing in ν. When $r = 0$, the result follows from that for V. Given $\nu \sim N(w_s, \sigma^2/s)$ and $\nu' \sim N(w'_s, \sigma^2/s)$ with $w_s \leq w'_s$, we have $V(\nu, m, n, 0) \leq V(\nu', m, n, 0)$ for any m and n, and

$$m_\nu = M(\nu, n, 0) = V(\nu, m_\nu, n, 0) \leq V(\nu', m_\nu, n, 0) \leq \sup\{m : V(\nu', m, n, 0) = c_f + m\} = M(\nu', n, 0) = m_{\nu'}. \ \square$$

Proof of Theorem 3. Recall from the proof of Theorem 2 that $\widetilde{\Lambda} = \inf\{n \in \mathbb{N} : HV(\nu_n, m_0, n) = c_f + m_0\}$ and that $\widetilde{\Lambda} \overset{d}{=} 1 + \Lambda_i(\pi^*)$. Since $q_n = q$ for all n, the proof of this result hinges on showing that

$$\mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^{t+1} q\right] = \mathbb{E}\left[\gamma^{\widetilde{\Lambda}}\mathbb{1}(\widetilde{\Lambda} = \Lambda)\right], \tag{27}$$

that would imply that

$$\mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^t c(Z(\nu_t, t)) + \gamma^{\widetilde{\Lambda}}(c_q - c_f)\mathbb{1}(\widetilde{\Lambda} = \Lambda)\right] = \mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^t \{c(Z(\nu_t, t)) + \gamma q(c_q - c_f)\}\right], \tag{28}$$

and this gives us an alternative representation for $C_{\pi^*}(\nu_0)$. We know that under the optimal selection policy

$$C_{\pi^*}(\nu_0) = \mathbb{E}\left[\sum_{k=1}^{\infty} \gamma^{\sum_{j=0}^{k-1} \widetilde{\Lambda}_j} \sum_{t=0}^{\widetilde{\Lambda}_k - 1} \gamma^t \{c_h \mathbb{1}(t = 0) + c(Z(\nu_{\pi(t)}, n_{\pi(t)}))\} + \gamma^{\widetilde{\Lambda}_k}(c_q - c_f)\mathbb{1}(\widetilde{\Lambda}_k = \Lambda_k) + \gamma^{\widetilde{\Lambda}_k} c_f\right]$$

where the $\{\widetilde{\Lambda}_j\}_{j=1}^{\infty}$ is the *iid*-sequence of optimal working lifetime of employees and $\widetilde{\Lambda}_j \overset{d}{=} \widetilde{\Lambda}$ for all j. Then, (28) allows us to write $C_{\pi^*}(\nu_0)$ as

$$C_{\pi^*}(\nu_0) = \mathbb{E}\left[\sum_{k=1}^{\infty} \gamma^{\sum_{j=0}^{k-1} \widetilde{\Lambda}_j} \sum_{t=0}^{\widetilde{\Lambda}_k - 1} \gamma^t \left\{ c_h \mathbb{1}(t=0) + c(Z(\nu_{\pi(t)}, n_{\pi(t)})) + \gamma q(c_q - c_f)\right\} + \gamma^{\widetilde{\Lambda}_k} c_f \right]$$

$$= \inf_{\pi \in \Pi} \left\{ c_h + \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \left\{ c\left(Z(\nu_{\pi(t)}, n_{\pi(t)})\right) + \gamma q(c_q - c_f) + (c_f + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1))\right\}\right]\right\}. \quad (29)$$

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

$$C_{\pi^*}(\nu_0) = \inf_{\pi \in \Pi} \left\{ c_h + \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t c\left(Z(\nu_{\pi(t)}, n_{\pi(t)})\right) + (c_f + c_h)\mathbb{1}(\pi(t) \neq \pi(t-1))\right]\right\} + \frac{\gamma q(c_q - c_f)}{1 - \gamma}, \quad (30)$$

and the solution to the minimization problem on the right hand side is the same as the solution to that minimization problem if the hiring cost is $c_h + c_f$ and the firing and quitting costs are set equal to 0. As a consequence $M_i(\nu_{i,n_i}, n_i, 0, c_h, c_f, c_q) < m_0(c_h, c_f, c_q)$ if and only if $M_i(\nu_{i,n_i}, n_i, 0, c_h + c_f, 0, 0) < m_0(c_h + c_f, 0, 0)$.

To complete our argument we then need to prove (27). The left-hand side satisfies

$$\mathbb{E}\left[\sum_{t=0}^{\widetilde{\Lambda}-1} \gamma^{t+1} q\right] = \sum_{t=1}^{\infty} \gamma^t q \mathbb{P}(\widetilde{\Lambda} \geq t) \quad (31)$$

and that the right-hand side satisfies

$$\mathbb{E}\left[\gamma^{\widetilde{\Lambda}} \mathbb{1}(\widetilde{\Lambda} = \Lambda)\right] = \sum_{t=1}^{\infty} \gamma^t \mathbb{P}(\Lambda = t, \widetilde{\Lambda} = t). \quad (32)$$

Now note that by using the fact that $\widetilde{\Lambda} \overset{d}{=} 1 + \Lambda_i(\pi^*)$ and the definition of q in (4) we have

$$q\mathbb{P}(\widetilde{\Lambda} \geq t) = \mathbb{P}(L_{t-1} = 1 | \Lambda(\pi^*) \geq t-1)\mathbb{P}(\widetilde{\Lambda} \geq t) = \mathbb{P}(L_{t-1} = 1 | \widetilde{\Lambda} \geq t)\mathbb{P}(\widetilde{\Lambda} \geq t) = \mathbb{P}(L_{t-1} = 1, \widetilde{\Lambda} \geq t)$$

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