Pyramids, Diamonds and Oscillations: What AI Does to Talent Management, Worker Careers, and Future Structure of Internal Labor Markets*

Guido Friebel Yao Huang Jin Li Yunchou Zhang

February 12, 2026

Abstract

AI is changing workplaces: it leads to higher individual productivity and faster learning. We model the effects of these changes on internal labor markets. We show that for productivity shocks, firms preserve the span of control in the long run: a pyramid remains a pyramid. In the short run, however, firms freeze junior hiring, temporarily transforming the pyramid toward a diamond. In the transition to the new steady state, firms may oscillate between pyramids and diamonds before settling. For learning shocks, the long-run span of control falls, potentially permanently shifting the firm from a pyramid to a diamond. These fluctuations create inequality between junior cohorts and affect firm value when human capital is firm-specific.

Keywords: Span of control, knowledge work, labor markets, firm-specific human capital

*Contact information: Guido Friebel (gfriebel@wiwi.uni-frankfurt.de), Department of Management and Microeconomics, Goethe University Frankfurt, 60323 Frankfurt/Main, Germany; Yao Huang (yaohuang@connect.hku.hk), Jin Li (jli1@hku.hk), Yunchou Zhang (zyc616@connect.hku.hk), Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong SAR.

1 Introduction

On-the-job training and learning by doing are among the most important determinants of human capital accumulation. As Arrow (1962) argued, learning is the product of experience and can only take place through the attempt to solve problems. While schools and universities supply conceptual foundations, workers develop expertise by doing. The internal labor markets of firms (Doeringer and Piore, 1971; Baker, Gibbs and Holmstrom, 1994*a*) provide the structure in which this progression takes place: junior workers build human capital through the tasks they perform on the job, and they gradually advance to more important positions (Gibbons and Waldman, 1999*a*, 2004; Waldman, 2013).

As AI automates many of the tasks that junior workers have traditionally performed, such as collecting data and drafting routine documents, firms across industries are starting to hire fewer entry-level workers[1]. This trend suggests that the traditional pyramid is giving way to a diamond: fewer juniors at the base, a bulge of experienced workers in the middle. PwC's chief AI officer predicts that "human-AI collaboration could boost productivity and speed by 50 percent" and that "what emerges will be a shift of the traditional labor pyramid to a diamond shape" (Priest, 2025).

Yet, if the tasks that once served as the bottom rungs of the career ladder are disappearing, what happens to the pipeline that produces tomorrow's senior workers? Forward-looking executives have warned that cutting junior hiring is dangerously short-sighted. As AWS CEO Matt Garman put it, the idea of firing juniors because AI can do their jobs is "the dumbest thing I've ever heard…how's that going to work when ten years in the future you have no one that has learned anything" (Sharwood, 2025).

At the heart of the matter lies a single tension: junior positions serve a dual purpose. In the short run, juniors perform tasks and generate output. In the long run, they are the training ground through which the firm replenishes its senior workforce. A firm that cuts junior hiring captures immediate efficiency gains but may be undermining its own future.

The reduction in junior hiring thus raises a set of interconnected questions that go well beyond the immediate cost savings. What happens to the structure of the internal labor market: does the pyramid permanently become a diamond, or is this a transient phenomenon? What happens to the accumulation of human capital in the economy when fewer workers pass through the entry-level positions where skills are built? How are

[1] Brynjolfsson, Chandar and Chen (2025) document a 16 percent decline in early-career hiring in AI-exposed fields, while demand for senior workers remains stable. Lichtinger and Hosseini Maasoum (2025), using résumé data covering 62 million U.S. workers across 285,000 firms, show that generative AI adoption sharply reduced junior employment. Similar patterns appear in freelance markets (Hui, Reshef and Zhou, 2024)

workers' careers and wages affected? And what happens to the firm's own profitability over time, especially when experienced workers carry knowledge that cannot easily be replaced from outside?

In this paper, we develop a model of internal labor markets to address these questions. A CEO allocates a fixed set of tasks to senior and junior workers, where seniors are more productive because they have accumulated human capital. Juniors learn on the job and, with some probability, acquire the skills needed to become seniors. The firm must therefore manage the intertemporal tradeoff between current production and the future supply of experienced workers. The setup is deliberately simple, but it captures the essential features of pyramidal organizations in which cohorts of entry-level workers sustain a talent pipeline into senior ranks. A key feature of the model is that the ratio of juniors to seniors is determined endogenously rather than fixed at one-to-one. This flexibility is what enables us to study the shape of the internal labor market and how that shape responds to AI.

We distinguish three channels through which AI may affect the workplace: it can raise the productivity of senior workers, raise the productivity of junior workers, or accelerate the rate at which juniors learn and qualify for senior positions. These are distinct shocks, and they lead to different answers.

Consider first the case in which AI increases productivity—whether of seniors or juniors. In the long run, the span of control, the ratio of juniors to seniors, does not change. A pyramid remains a pyramid. The firm shrinks, because fewer workers of both types are needed to meet the same production target, but the organizational shape is preserved. The wage ratio between seniors and juniors does shift, however, and its direction depends on whose productivity increases: when seniors become more productive, the wage gap widens; when juniors become more productive, it narrows.

The short run, however, looks very different from the long run. Immediately following the shock, firms hire fewer juniors or even fire them; the pyramid temporarily contracts toward a diamond. The workforce then gradually recovers toward the new, smaller steady state. But this recovery need not be smooth. Under certain conditions, the firm oscillates: the initial hiring freeze starves the talent pipeline, creating a future shortage of seniors, which forces a surge in junior hiring, which in turn produces an oversupply of seniors, and so on. The firm swings between pyramid and diamond before finally settling. These oscillations are more likely when the productivity gap between seniors and juniors is large, when learning is rapid, and when senior attrition is high.

The picture changes when AI primarily accelerates learning. If juniors acquire skills faster, the long-run span of control permanently falls: fewer juniors are needed to sustain the same flow into the senior ranks, and the firm shifts from a pyramid to a diamond.

This is the only type of AI shock that permanently reshapes the organizational hierarchy. The transition, however, can again involve oscillations when the acceleration in learning is sufficiently large.

Finally, we extend the model to allow for firm-specific human capital, where internally promoted seniors are more productive than external hires. In this setting, firms hire exclusively from within, which makes the talent pipeline even more consequential: when it breaks down, the firm cannot simply plug the gap from the outside market, and its value falls. An additional finding emerges for the wage structure. When AI accelerates learning, the wage ratio between seniors and juniors compresses. Faster learning raises the value of the junior layer—each junior now generates firm-specific capital more quickly—which narrows the premium that seniors command.

These results have implications for management. While the forward-looking managers in our model contract the junior layer to capture immediate efficiency gains, it can disrupt the firm's internal supply of talent and reduce the firm's future value. Therefore, when the stability of future talent supply is important, a CEO may need to smooth these fluctuations by managing the talent pipeline.

Our results also have implications for inequality between junior cohorts. For different cohorts of juniors, the employment and learning opportunities can vary significantly. Juniors entering during a period of excess seniors face reduced opportunities to be employed and acquire skills. [Matt Beane Skill Codes: "There is a whole generation of lawyers missing out on training and professional development."] This "lost cohort," however, may also create a shortage of seniors in the future. This shortage benefits the next cohort by increasing their opportunities to work and learn.

The ingredients of the model are inspired by a large literature on internal labor markets, a concept first introduced by Doeringer and Piore (1971). Their seminal work triggered a thriving theoretical and empirical literature, summarized by Gibbons (1997), Gibbons and Waldman (1999*a*), Lazear and Oyer (2013), and Waldman (2013).

A number of empirical papers, most notably, Baker, Gibbs and Holmstrom (1994*a*,*b*) provide a detailed case study of a firm indicating that ILM were conducive to human capital acquisition and providing promotion incentives (that are absent in our paper).[2] Pastorino (2024) investigates the importance of human capital accumulation and learning about employees' abilities, especially for sorting employees to jobs, within one firm. Recent contributions underline the importance of internal labor markets and offer estimates that a

[2]Our work abstracts away from the incentive role of promotions, which are an important topic in ILM (Lazear and Rosen, 1981; Rosen, 1986; Malcomson, 1984; MacLeod and Malcomson, 1988; Prendergast, 1993; Gibbons and Waldman, 1999*b*; Zabojnik and Bernhardt, 2001; Waldman, 2003; Kräkel and Schöttner, 2012; Auriol, Friebel and Von Bieberstein, 2016; Ke, Li and Powell, 2018; Bianchi et al., 2023).

large proportion of workers are covered by ILM practices despite increasing competitive pressure (Huitfeldt et al., 2023; Osterman, 2024).

Our work is related to the literature regarding vacancy chains and slot constraints in internal labor markets (Simon, 1951; White, 1970; Beckmann, 1978; Stewman and Konda, 1983; Rosenbaum, 1984; Ke, Li and Powell, 2018). These structural features impose constraints that bind the careers of workers across different levels. This interdependence creates career spillovers, where the progression of junior workers relies on the departure of seniors, a phenomenon documented empirically by Friebel and Panova (2008); Bianchi et al. (2023).

How AI affects the development of human capital has recently been modelled by Ide and Talamàs (2025) and Garicano and Rayo (2025). Both papers adapt the apprenticeship idea first brought and modeled by Garicano and Rayo (2017) and Fudenberg and Rayo (2019). Junior people develop human capital through their contact with experts or an explicit intertemporal transfer of knowledge from a senior to a junior. We believe that it is important to bring firms explicitly into the picture in order to better understand the dynamics of junior and senior cohorts in the long run and in transitions. Profit-maximizing firms manage the intertemporal effects of hiring juniors today who can become seniors tomorrow. This has implications for the overall supply of human capital in an economy and does not only affect the intensive margin of how much human capital each worker has, but also how many trained juniors the economy has who can, later in their life cycles take on more complex tasks.

Our production function is one in which CEOs need to get N tasks done by delegating to seniors and juniors depending on their productivity, which is enhanced by AI. This is simplistic and we can augment the setting. It should be noted though that this production function is grounded in the emerging empirical literature that shows that more tasks can be done by AI in a given time and leads to faster learning. Generative AI significantly increases worker productivity, particularly for novice and lower-skilled individuals, allowing less experienced workers to rapidly close performance gaps with top performers (Noy and Zhang, 2023; Brynjolfsson, Li and Raymond, 2025; Chen et al., 2025). This internal augmentation may translate into external displacement; as AI automates entry-level tasks, firms may be freezing hiring for junior roles. Beyond labor demand, reliance on AI introduces risks to human capital formation and creative diversity. While AI tools can help workers move down the experience curve faster in supported environments (Brynjolfsson, Li and Raymond, 2025), unfettered access in educational settings can act as a "crutch," improving immediate performance but significantly degrading unassisted learning and exam scores once the tool is removed (Bastani et al., 2025).

2 Model Setup

Consider an economy populated by a continuum of CEOs with total mass M. Each CEO organizes a firm by hiring senior and junior workers to complete a fixed target of N tasks. As each CEO defines a firm, we use the terms "CEO" and "firm" interchangeably. Time is discrete and indexed by $t = 0, 1, 2, \cdots, \infty$.

Workers There is an infinite supply of junior workers with outside option $\underline{v} > 0$. In each period, junior workers who finish tasks acquire necessary skills and become seniors with probability $q \in (0, 1]$. At the end of each period, junior and senior workers exit the market exogenously at rates $d_l \in (0, 1)$ and $d_h \in (0, 1)$, obtaining the same outside option $\underline{v}$. Let $w_t^l > 0$ and $w_t^h > 0$ denote the wages paid to juniors and seniors. Let v_t^l and v_t^h denote the value functions for employed juniors[3] and seniors:

$$v_t^l = w_t^l + \delta(1 - d_l)[qv_{t+1}^h + (1 - q)v_{t+1}^l] + \delta d_l \underline{v}, \tag{VF^l}$$

$$v_t^h = w_t^h + \delta(1 - d_h)v_{t+1}^h + \delta d_h \underline{v}. \tag{VF^h}$$

Because of the infinite supply of juniors, free entry drives the value of being an employed junior down to the outside option: $v_t^l = \underline{v}$.

Firm In every period, firm i hires $n_{i,t}^l$ juniors and $n_{i,t}^h$ seniors to complete a fixed target of N tasks, where $n_{i,t}^l \geq 0$ and $n_{i,t}^h \geq 0$. Every task generates revenue of y. Each junior worker can finish γ^l tasks, and each senior worker can finish γ^h tasks, where $0 < \gamma^l < \gamma^h$. Let $D_{i,t}^l \geq 0$ and $D_{i,t}^h \geq 0$ denote the tasks allocated to juniors and seniors. Then we have the following capacity constraints:

$$\begin{aligned} \gamma^l n_{i,t}^l &\geq D_{i,t}^l, \\ \gamma^h n_{i,t}^h &\geq D_{i,t}^h, \\ D_{i,t}^l + D_{i,t}^h &\leq N. \end{aligned} \tag{CC}$$

These capacity constraints show that each firm has limited tasks. If any firm hires more workers, its real output is still N tasks. The firm gets the following per-period payoff:

$$\pi_{i,t} = (D_{i,t}^l + D_{i,t}^h)y - w_t^h n_{i,t}^h - w_t^l n_{i,t}^l.$$

[3] Here we focus on the juniors allocated tasks. We will show in Lemma 1 that all juniors will participate in production given the market wage w_t^l.

Firms are forward-looking and maximize the present value of discounted profits. Hence, the optimization problem for firm i is:

$$\max_{\{n^l_{i,t}, n^h_{i,t}, D^l_{i,t}, D^h_{i,t}\}^\infty_{t=0}} \quad \sum_{t=0}^{\infty} \delta^t \left[(D^l_{i,t} + D^h_{i,t}) y - w^h_t n^h_{i,t} - w^l_t n^l_{i,t} \right],$$

subject to the capacity constraints (CC) for all t.

Market Clearing Let N^h_t denote the aggregate supply of senior workers. Let $N^l_t := \int_0^M n^l_{i,t} di$ denote the aggregate number of junior workers who participate in production.[4] The aggregate supply of senior workers is affected by the human capital accumulation in the last period:

$$N^h_{t+1} = (1 - d_h) N^h_t + q(1 - d_l) N^l_t. \tag{FC}$$

Market clearing requires that aggregate demand of seniors from firms equals the aggregate supply:

$$\int_0^M n^h_{i,t} di = N^h_t.$$

Equilibrium A competitive equilibrium consists of wage sequences $\{w^l_t, w^h_t\}^\infty_{t=0}$, firm hiring plans $\{n^l_{i,t}, n^h_{i,t}\}^\infty_{t=0}$, and aggregate labor stocks $\{N^l_t, N^h_t\}^\infty_{t=0}$, such that: Given wages, firm hiring plans maximize the present value of discounted profits; Value functions satisfy the Bellman equations (VF^l and VF^h), and the free-entry condition holds; Aggregate demand equals aggregate supply, which evolves according to the flow equation (FC).

Timing Figure 1 shows the timeline of the stage game. At the beginning of period t, firms decide whether they should hire or fire workers. Then firms allocate the tasks between senior and junior workers. After that, the production happens and the output is realized. During the production process, junior workers acquire skills with certain probability. Finally, some workers leave, and those remaining juniors who have acquired skills turn into seniors and get promoted.

Assumption 1. $\gamma^l y > (1 - \delta)\underline{v}$.

Assumption 1 ensures that the value of output is so high that it is always optimal to have all the tasks finished.

[4] Similar to the value function, here we also implicitly assume that there is no idle junior worker.

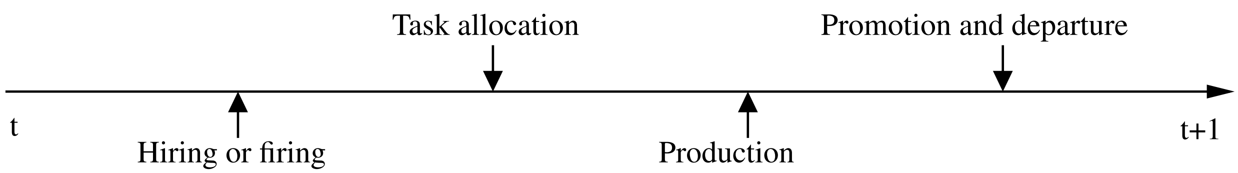


Figure 1: Stage Game Timeline

3 Steady State Analysis

To study the steady state, we first investigate the optimal task allocation and hiring policy for each firm. The access to a competitive labor market allows firms to adjust their workforce size and composition immediately. This flexibility decouples the firm's decisions across periods so that a hiring choice in period t does not affect the firm in period $t+1$. Therefore, given the wages in each period, the firm's optimization problem can be reduced to a sequence of static problems:

$$\max_{n^l_{i,t}, n^h_{i,t}, D^l_{i,t}, D^h_{i,t}} \quad (D^l_{i,t} + D^h_{i,t})y - w^h_t n^h_{i,t} - w^l_t n^l_{i,t},$$

subject to the capacity constraints (CC).

After solving the optimization problem, we characterize the task allocation and hiring policy in the following lemma.

Lemma 1. *In a competitive market, the following hold for firm i: (i) $D^l_{i,t} + D^h_{i,t} = N$; (ii) $D^l_{i,t} = \gamma^l n^l_{i,t}$; (iii) $D^h_{i,t} = \gamma^h n^h_{i,t}$.*

Lemma 1 shows that all the capacity constraints (CC) bind in a competitive market. It is natural that all tasks should be allocated and finished as we assume that the output of the task is sufficiently high. With infinite supply of junior workers, every firm should have enough workers to finish the tasks. There is no gain from having excessive junior workers. Juniors only pick up skills from working, meaning that firms should not hire juniors without allocating tasks to them. In a competitive market, a single firm cannot affect the market wage for seniors. Therefore, firms will not hold some seniors idle to help more juniors build up human capital.

The following lemma characterizes the wage ratio between senior and junior workers.

Lemma 2. *If firms employ strictly positive quantities of both senior and junior workers ($n^h_{i,t} > 0$ and $n^l_{i,t} > 0$), the ratio of wages must equal the ratio of productivities:*

$$\frac{w^h_t}{w^l_t} = \frac{\gamma^h}{\gamma^l}.$$

Lemma 2 shows a no-arbitrage condition. The firm's production technology is linear, making junior and senior workers perfect substitutes. This linearity imposes an equilibrium condition on wages for firms to hire both types of workers. Because the firm can instantly substitute γ^h/γ^l junior workers for one senior worker, the market wage premium for seniors must exactly match their productivity premium.

Now we can investigate the steady state. In particular, all stocks and flows are constant over time in a steady state. Given the linear production technology and competitive labor markets, for individual firms, any combination of seniors and juniors satisfying the binding capacity constraints is optimal. For simplicity, we focus on the symmetric equilibrium, where each firm adopts the same hiring strategy and mirrors the aggregate economy.[5] Let n^{l*} and n^{h*} denote the number of junior and senior workers for a representative firm in a steady state. We have the following steady state constraints:

$$\begin{aligned} n^h_{t+1} &= n^h_t = n^{h*}; \\ n^l_{t+1} &= n^l_t = n^{l*}. \end{aligned} \tag{SS}$$

Let $s_t := n^l_t/n^h_t$ denote the span of control in period t, and s^* represents that in a steady state. Then we have the following lemma characterizing the steady state.

Lemma 3. *The steady state of a symmetric equilibrium is characterized as follows:*

(i) *The equilibrium span of control and employment levels are:*

$$s^* = \frac{d_h}{q(1-d_l)}, \quad n^{l*} = \frac{s^*}{\gamma^l s^* + \gamma^h} N \quad \textit{and} \quad n^{h*} = \frac{1}{\gamma^l s^* + \gamma^h} N.$$

(ii) *The equilibrium wages are:*

$$w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)}\right] \quad \textit{and} \quad w^{h*} = \frac{\gamma^h}{\gamma^l} w^{l*}.$$

Part (i) of Lemma 3 characterizes the firm structure and the number of workers in each layer in the steady state. It shows that the flow rates determine the span of control (s^*), the ratio of juniors to seniors. This ratio, together with productivity (γ^l and γ^h) and the total number of tasks (N), determines the size of each layer (n^{l*} and n^{h*}).

[5] While the current framework allows for asymmetric distributions of workers across firms, if internal hiring of seniors has firm-specific human capital, the symmetric stationary equilibrium is the unique stationary equilibrium. We show this in Section 5.

The expression of s^* shows that the productivity does not affect the span of control in the steady state. Only the flow rates (i.e., learning rate q and departure rates d_h and d_l) matter. The firm looks like a pyramid $(s^* > 1)$ when seniors leave often (high d_h), juniors leave often (high d_l), or juniors learn slowly (low q). In this case, the firm must hire many juniors just to keep the senior layer staffed. Conversely, if seniors stay longer (low d_h), juniors stay longer (low d_l), or juniors learn faster (high q), then firm needs fewer juniors to feed the senior layer. That is, the structure becomes more like a diamond $(s^* < 1)$.

Note that once the flow rates determine the span of control, the number of tasks (N) and workers' productivity (γ^h and γ^l) determine the size of each layer. The firm adjusts the structure until the combined output of the two layers meets N.

Part (ii) characterizes the steady-state wages. The junior wage w^l equals the flow outside option, $(1-\delta)\underline{v}$, multiplied by a factor strictly less than one. This is because free entry drives the value of a junior position to $\underline{v}$, and workers accept lower current pay in exchange for the continuation value of acquiring skills and earning the senior wage w^h. The senior wage is then determined by the no-arbitrage condition in Lemma 2 $(w^h/w^l = \gamma^h/\gamma^l)$.

How does AI change the firm structure? We distinguish two effects AI may have. First, AI may be a *productivity shock* that increases workers' capacity to take on tasks, γ^h or γ^l. Second, AI may be a *learning shock* that increases the probability q that a junior worker acquires high ability. Now we summarize how AI shocks affect the firm structure in the long run.

Proposition 1. *The steady state reacts as follows to the AI shock:*

(i) *If senior productivity γ^h increases, the number of workers drops in both layers (n^{h*} and n^{l*}), but the span of control (s^*) remains constant. The junior wage (w^{l*}) decreases, while the senior wage (w^{h*}) increases.*

(ii) *If junior productivity γ^l increases, the number of workers drops in both layers (n^{h*} and n^{l*}), but the span of control (s^*) remains constant. The junior wage (w^{l*}) increases, while the senior wage (w^{h*}) decreases.*

(iii) *If the learning probability q increases, the number of seniors (n^{h*}) increases, the number of juniors (n^{l*}) decreases, and the span of control (s^*) decreases. Both the junior wage (w^l) and the senior wage (w^{h*}) decrease, but the ratio ($\frac{w^{h*}}{w^{l*}}$) stays constant.*

Proposition 1 distinguishes how different AI shocks shape the internal labor market. First, consider AI as a productivity shock, i.e., an increase in γ^h or γ^l. This shock reduces the total workforce required to meet the production target N. While the total workforce shrinks, the reduction occurs uniformly throughout the hierarchy. This is because the span of control (s^*) depends only on the flow rates (d_h, d_l, q), not on worker productivity. Consequently, the organization shrinks in workforce but retains its structure in the long run; a pyramidal firm will shrink in size but still maintain its pyramidal structure.

The impact of productivity shock on wages depends on which type of workers are affected. When seniors become more productive (γ^h increases), the productivity gap between seniors and juniors widens, leading to a rise in senior wage. This, in turn, increases the continuation value of a junior position. Because of the higher continuation value and the free entry of juniors, firms offer lower junior wages. In contrast, when juniors become more productive (γ^l increases), the productivity gap narrows. This decreases senior wage and compresses the continuation value of a junior position. Hence, firms will have to compensate juniors with a higher wage.

Second, consider AI as a learning shock, i.e., an increase in q. A higher q raises the probability that a junior worker qualifies for promotion. This higher probability increases the flow of workers into the senior layer, increasing the number of seniors. As the flow rate upward increases, the firm needs fewer juniors to ensure the steady flow of talents. This reduces the span of control s^*: a pyramidal firm will be more likely to adopt a diamond structure in the long run.

The effect of learning shock on wages shows how faster learning decreases the scarcity of seniors. If AI accelerates learning (q increases), it increases the continuation value of a junior position. Because of free entry, juniors accept lower wages (w_l) today. As juniors become cheaper, the senior wage (w_h) must also decrease to satisfy the no-arbitrage condition.

4 AI Shocks and Transition Dynamics

In this section, we analyze how firms adjust the firm structure over time following an AI shock. We first discuss the flow of senior workers, which determines the transition path between steady states. We then show how productivity and learning shocks affect this path.

4.1 The Supply of Seniors

The transition of the firm structure depends on the transition of its senior workforce. By substituting the optimal hiring policy from Lemma 1 into the flow condition, we obtain a map linking the current stock of seniors, n_t^h, to the future stock, n_{t+1}^h. We call this map the supply of seniors S:

$$n_{t+1}^h = S(n_t^h) \coloneqq (1-d_h)n_t^h + q(1-d_l)n^l(n_t^h).$$

The supply $S(n_t^h)$ determines how many seniors will be available in the next period based on how many seniors are employed now. The supply consists of two opposing forces: retention and the pipeline.

Retention, $R(n_t^h)$, represents the incumbents who remain with the firm:

$$R(n_t^h) = (1-d_h)n_t^h.$$

Retention is strictly increasing: more seniors today imply more seniors tomorrow.

The pipeline, $P(n_t^h)$, represents junior workers who acquire skills and become seniors:

$$P(n_t^h) = q(1-d_l)n^l(n_t^h) = \begin{cases} \frac{q(1-d_l)}{\gamma^l}(N-\gamma^h n_t^h) & \text{if } n_t^h < N/\gamma^h, \\ 0 & \text{if } n_t^h \geq N/\gamma^h. \end{cases}.$$

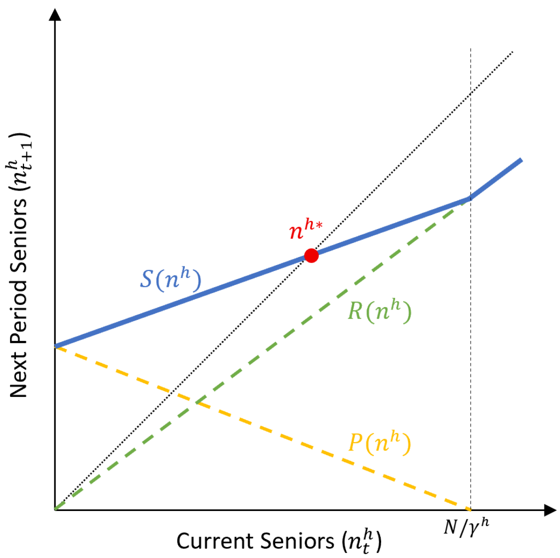
Unlike retention, the pipeline decreases in n_t^h. Because the firm hires juniors to complete tasks that seniors cannot, a larger senior workforce crowds out juniors. This crowding out shrinks the pool of future seniors. Once the senior workforce meets the production target ($n_t^h \geq N/\gamma^h$), the firm freezes junior hiring, reducing the pipeline to zero.

Together, retention and the pipeline determine the supply of seniors, $S(n_t^h)$. Figure 2 shows this relationship. If retention dominates the crowding-out effect of the pipeline, more seniors today means more seniors tomorrow ($S' > 0$). If the pipeline dominates retention, more seniors today means fewer seniors tomorrow ($S' < 0$). As we show in the rest of this section, the relative strength of these two forces affects the firm's transition path.

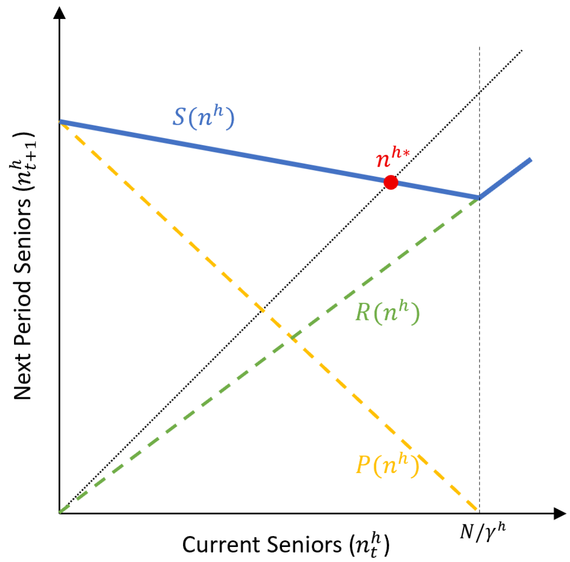
Before we explore the transitions in detail, note that the steady state occurs at the intersection of $S(n_t^h)$ and the 45-degree line (Figure 2). To ensure the firm converges to a steady state, we assume the system is stable ($|S'| < 1$).[6] We also note that while the workforce adjusts slowly according to S, wages adjust instantly to their new steady-state levels.[7]

[6]If the system is unstable ($S' \leq -1$), the labor market can involve permanent cycles between oversupply and scarcity of seniors.

[7]This is driven by the no-arbitrage condition and value functions of workers.



(a) Retention Dominates (Positive Slope)



(b) Pipeline Dominates (Negative Slope)

Figure 2: Retention, Pipeline, and Supply of Seniors

4.2 AI as a Productivity Shock to Seniors

Suppose AI increases the productivity of senior workers from γ_0^h to γ_{AI}^h at the beginning of period $t = 1$. The following proposition characterizes the transition.

Proposition 2. *Assume the AI shock is not so extreme that the firm still hires juniors at $t = 1$ ($\gamma_{AI}^h < N/n_0^{h*}$). Following an increase in senior productivity from γ_0^h to γ_{AI}^h at $t = 1$:*

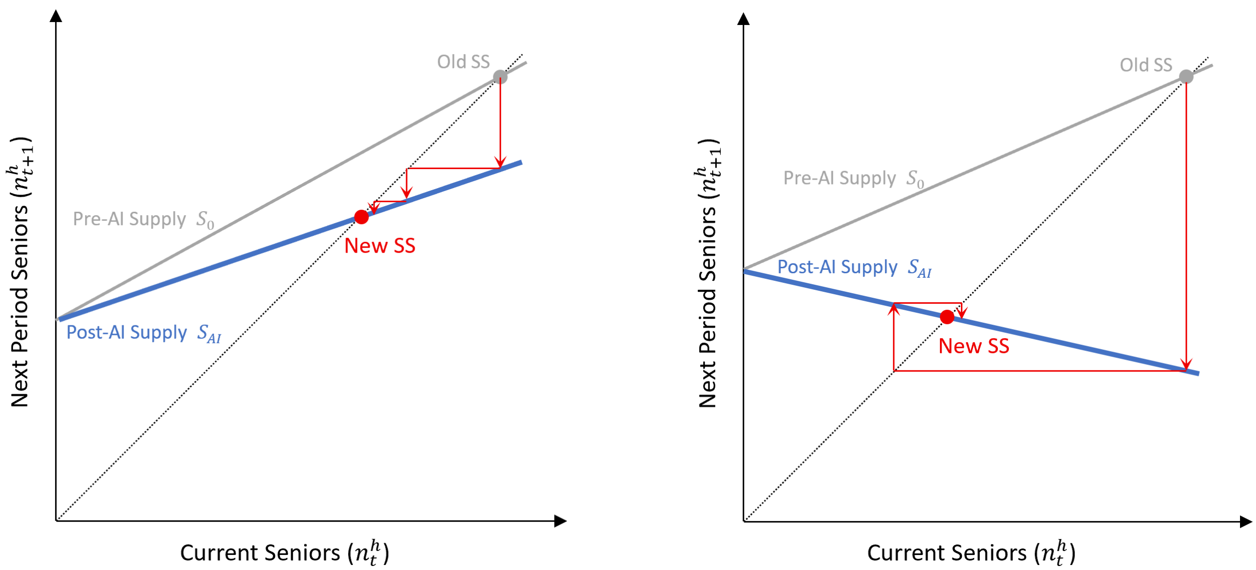
(*i*) *At $t = 1$, the senior workforce remains at n_0^{h*}, the junior workforce drops to $n_1^l < n_0^{l*}$, and the span of control decreases.*

(*ii*) *For $t \geq 2$, the firm structure converges toward the new steady state according to the following cases:*

  (*a*) *If $\gamma_{AI}^h \leq \frac{1-d_h}{q(1-d_l)}\gamma^l$, the transition is monotonic: the senior number n_t^h decreases and the junior number n_t^l increases. Consequently, the span of control n_t^l/n_t^h increases.*

  (*b*) *If $\gamma_{AI}^h > \frac{1-d_h}{q(1-d_l)}\gamma^l$, the transition oscillates: the senior number n_t^h alternates above and below its new steady state, and the junior number n_t^l moves in the opposite direction. Consequently, the span of control fluctuates.*

Proposition 2 highlights the contrast between the short-run and long-run effects of AI. In the long run, the shock does not alter the firm's span of control (Proposition 1): a pyramid remains a pyramid. In the short run, however, the span can change significantly: the pyramid may temporarily transform into a diamond.



(**a**) Small γ^h Increase (**b**) Large γ^h Increase

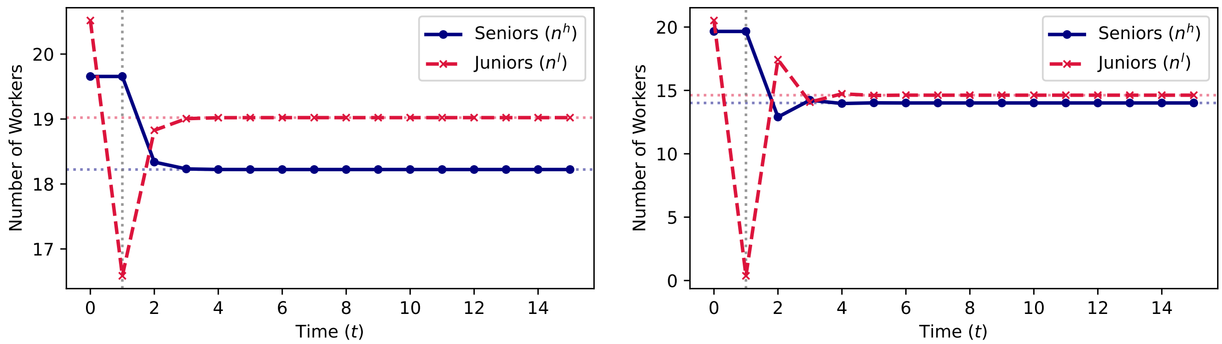
Figure 3: AI as a Productivity Shock to Seniors

As part (i) shows, the immediate impact of AI is a decrease in the span of control. This decrease is caused by a reduction in junior hiring. At $t = 1$, the firm holds a fixed stock of seniors ($n^h_1 = n^{h*}_0$) inherited from the previous period. Because each senior is now more productive, the firm requires fewer juniors to meet the production target N. Junior hiring therefore falls to $n^l_1 = (N - \gamma^h_{AI} n^{h*}_0)/\gamma^l$.[8] With the senior stock fixed and junior numbers falling, the span decreases; the firm can temporarily shift from a pyramid toward a diamond structure.

The drop in juniors at $t = 1$ then restricts the senior workforce at $t = 2$. The pipeline of new seniors tomorrow (P) depends on the juniors hired today. Because hiring of juniors dropped, the pipeline shrinks. That is, the increase in γ^h shifts the pipeline curve P and the supply curve S downward (see the change of S and the first transition arrow in Figure 3). With the pipeline choked, exiting incumbents (d_h) are not fully replaced, and the senior workforce begins to fall and converge to the new, lower steady state (see the second transition arrow to the left).

The subsequent path of this convergence depends on the magnitude of the AI-enabled productivity γ^h_{AI}, corresponding to the two cases identified in part (ii). A higher γ^h_{AI} amplifies the crowding-out effect (S' more likely to be negative). Intuitively, when seniors are more productive, each additional senior today crowds out more juniors, which starves the future pipeline more aggressively.

[8] To simplify the initial discussion, we assume the shock is not so extreme as to cause a total hiring freeze. We discuss extreme shocks in Appendix C.1.



(**a**) Smooth Transition: $\gamma^h_{AI} = 3.4$.

(**b**) Oscillations: $\gamma^h_{AI} = 5.05$.

Figure 4: Simulation of the Transition where AI Increases γ^h. Parameters: $d_l = 0.65$, $d_h = 0.35$, $q = 0.958$, $\gamma^h_0 = 3$, $\gamma^l = 2$, and $N = 100$.

If the new productivity γ^h_{AI} is small (case a and Figure 3a), the transition is smooth and monotonic. To see why, notice that if γ^h_{AI} is small, retention still dominates the crowding-out effect ($S' \geq 0$). In this case, although the initial excess seniors (compared to the new steady state) reduces junior hiring, it does not starve the pipeline enough to create a shortage. The senior workforce smoothly declines, while the junior workforce increases, toward the new steady state (Figure 4a). Therefore, the span of control increases.

Conversely, if the new productivity γ^h_{AI} is large (case b and Figure 3b), the transition oscillates. If γ^h_{AI} is large enough, crowding-out effect dominates retention ($S' < 0$). In this case, today's excess seniors crowd out so many juniors that the firm ends up with a senior shortage in the next period. This shortage then causes a large hiring of juniors, which in turn leads to excess seniors later. Therefore, the span of control oscillates: the firm structure can swing back and forth between pyramid and diamond before finally settling (Figure 4b).

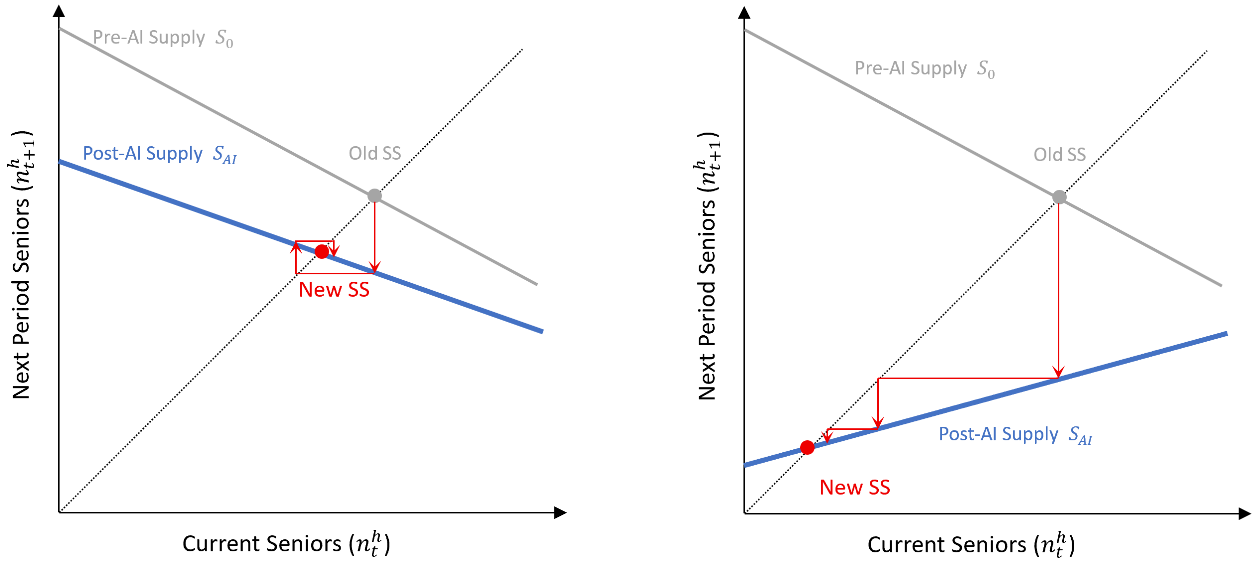
4.3 AI as a Productivity Shock to Juniors

Suppose AI increases the productivity of junior workers from γ^l_0 to γ^l_{AI} at the beginning of period $t = 1$. The following proposition characterizes the transition.

Proposition 3. *Following an increase in junior productivity from γ^l_0 to γ^l_{AI}:*

(*i*) *At $t = 1$, the senior workforce remains at n^{h*}_0, the junior workforce drops to $n^l_1 < n^{l*}_0$, and the span of control decreases.*

(*ii*) *For $t \geq 2$, the firm structure converges toward the new steady state according to the following cases:*



(**a**) Small γ^l Increase (**b**) Large γ^l Increase

Figure 5: AI as a Productivity Shock to Juniors

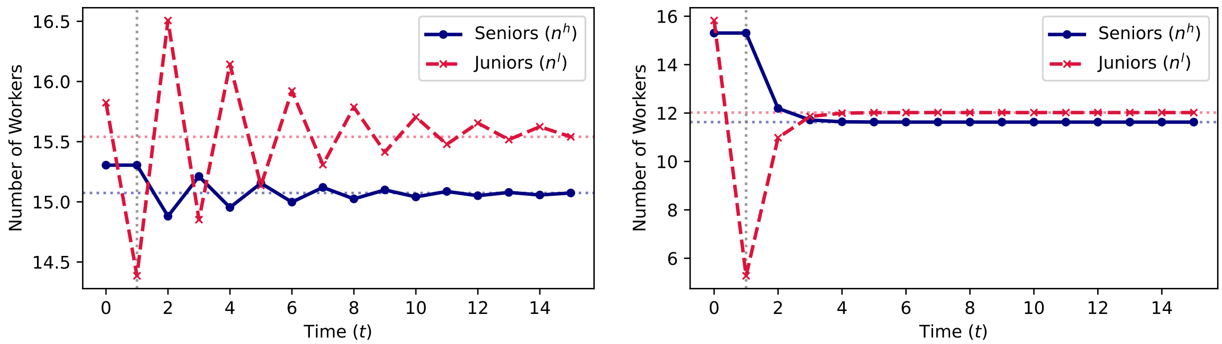
(*a*) *If* $\gamma^l_{AI} < \frac{q(1-d_l)}{1-d_h}\gamma^h$, *the transition oscillates: the senior number* n^h_t *alternates above and below its new steady state, and the junior number* n^l_t *moves in the opposite direction. Consequently, the span of control fluctuates.*

(*b*) *If* $\gamma^l_{AI} \geq \frac{q(1-d_l)}{1-d_h}\gamma^h$, *the transition is monotonic: the senior number* n^h_t *decreases and the junior number* n^l_t *increases. Consequently, the span of control* n^l_t/n^h_t *increases.*

Proposition 3 again shows the difference between the short-run and long-run effects of AI. Similar to the case with the senior productivity shock, the immediate impact at $t = 1$ is a sharp drop in the span of control (part i). Because each junior is now more capable, the firm requires fewer of them to complete the tasks that seniors cannot handle ($n^l_1 = (N - \gamma^h n^{h*}_0)/\gamma^l_{AI} < n^{l*}_0$). With the senior stock fixed, the firm again shifts temporarily toward a smaller span, or even a diamond structure.

The drop in juniors again restricts the pipeline (P) and shifts the supply curve S downward (see the change of S and the first transition in Figure 5). The senior workforce subsequently falls and converges to a new, lower steady state (see the second transition to the left).

While the senior workforce falls in both senior and junior productivity shocks, the transition paths differ. While an increase in γ^h destabilizes the transition and can cause oscillations, an increase in γ^l stabilizes it and facilitates a smooth transition. The intuition lies in the productivity gap γ^h/γ^l. When junior productivity rises, juniors become better substitutes for seniors, and the gap narrows. This implies that an additional senior dis-



(**a**) Oscillations: $\gamma^l_{AI} = 1.1$.

(**b**) Smooth Transition: $\gamma^l_{AI} = 3$.

Figure 6: Simulation of the Transition where AI Increases γ^l. Parameters: $d_l = 0.5$, $d_h = 0.305$, $q = 0.59$, $\gamma^h = 5.5$, $\gamma^l_0 = 1$, and $N = 100$.

places fewer juniors than before, which weakens the crowding-out effect (S' increases). Specifically, there are two cases.

If the new productivity γ^l_{AI} is small (case a and Figure 5a), the transition is oscillating. The reason is that, if γ^l_{AI} is still small, the crowding-out effect remains strong enough to dominate retention ($S' < 0$). The initial excess of seniors starves the pipeline, leading to the same oscillating pattern observed in the senior shock (see Figure 6a).[9]

If the new productivity γ^l_{AI} is large (case b and Figure 5b), the transition becomes smooth and monotonic. The crowding-out effect weakens and retention dominates ($S' \geq 0$). In this range, the initial excess of seniors causes only a mild reduction in the pipeline, not enough to create a shortage. The senior stock declines gradually and monotonically (see Figure 6b).

This result shows an asymmetry between the two types of productivity improvements. Improving senior productivity widens the gap between layers, strengthening crowding-out and volatility. Improving junior productivity compresses the gap, weakening crowding-out and smoothing the transition.

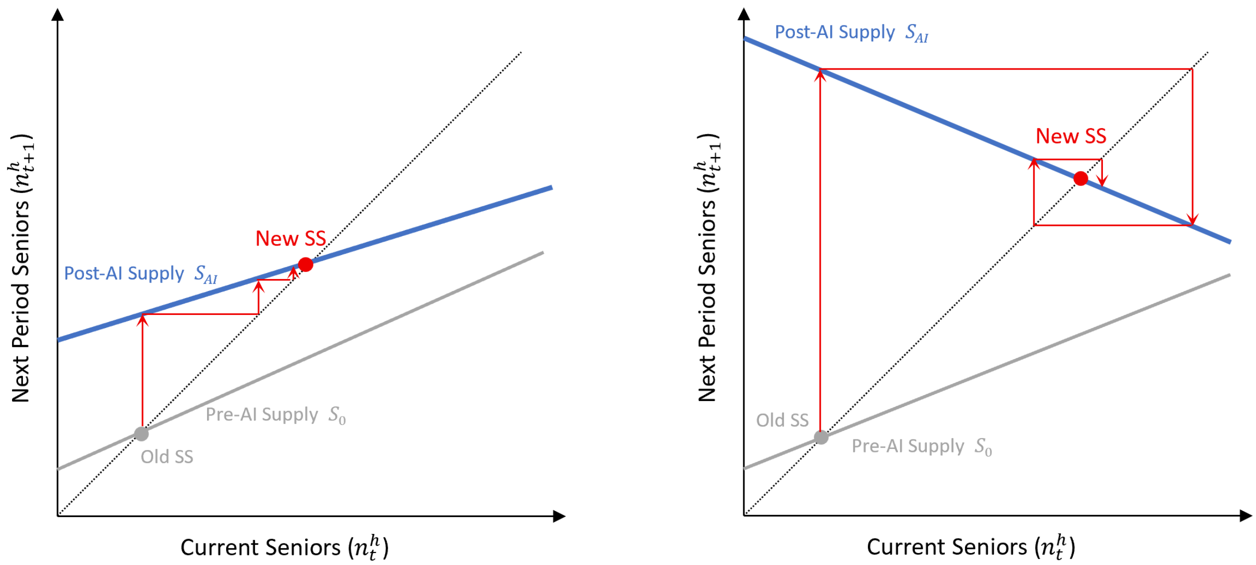
4.4 AI as a Learning Shock

Suppose AI increases the learning rate of junior workers from q_0 to q_{AI} at the beginning of period $t = 1$. The following proposition characterizes the transition.

Proposition 4. *Following an increase in the learning rate from q_0 to q_{AI}:*

(i) At $t = 1$, the senior and junior numbers remain at n^{h*}_0 and n^{l*}_0.

[9] One observation is that, in Figure 6a, not only the span of control in transition may be greater than pre-AI, but the absolute number of juniors hired may be greater than before (e.g., $t = 2$).



(**a**) Small q Increase   (**b**) Large q Increase

Figure 7: AI as a Learning Shock

(ii) *At $t = 2$, the senior number increases to $n^h_2 > n^{h*}_0$, while the junior number drops to $n^l_2 < n^{l*}_0$. Consequently, the span of control n^l_2/n^h_2 decreases.*

(iii) *For $t \geq 3$, the firm structure converges toward the new steady state according to the following cases:*

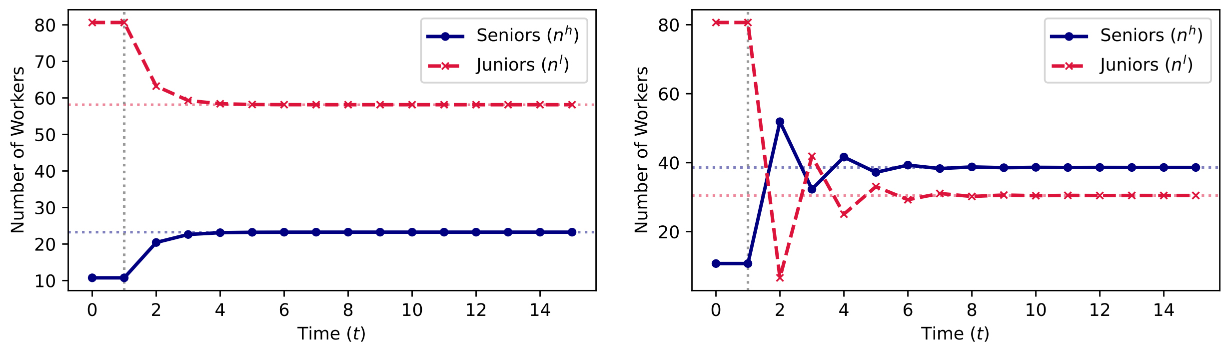
(a) *If $q_{AI} \leq \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, the transition is monotonic: the senior number n^h_t increases and the junior number n^l_t decreases. Consequently, the span of control n^l_t/n^h_t decreases.*

(b) *If $q_{AI} > \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, the transition oscillates: the senior number n^h_t alternates below and above its new steady state, and the junior number n^l_t moves in the opposite direction. Consequently, the span of control fluctuates.*

Unlike productivity shocks, a learning shock generates no change in the firm structure at $t = 1$ (part i). Since the productivity of workers (γ^h, γ^l) remains unchanged at $t = 1$, the firm requires the exact same number of juniors to supplement the existing seniors.

The shock hits the structure with a one-period lag (part ii). At $t = 2$, the higher learning probability q causes a surge of new seniors from the previous juniors. That is, the higher q shifts the pipeline (P) and therefore the supply curve (S) upward (Figure 7). The senior workforce grows, and the firm to hire fewer juniors to meet the production target. Consequently, the span of control decreases at $t = 2$.

The subsequent convergence (part iii) also depends on the magnitude of AI-enabled learning q_{AI}. A higher q strengthens the crowding-out effect. Intuitively, when q is high,



(**a**) Smooth Adjustment: $q_{AI} = 0.3$.

(**b**) Oscillations: $q_{AI} = 0.95$.

Figure 8: Transition Dynamics with Positive Learning Shock. Parameters: $d_l = 0.4$, $d_h = 0.45$, $q_0 = 0.1$, $\gamma^h = 1.8$, $\gamma^l = 1$, and $N = 100$.

every junior hired today represents a high probability of a new senior tomorrow. Consequently, any variation in junior hiring today translates into a large variation in the number of new seniors tomorrow. This amplifies the sensitivity of the pipeline and drives S' downward.

If the new learning probability q_{AI} is small (case a and Figure 7a), the transition is smooth and monotonic. If q_{AI} remains low, retention dominates ($S' \geq 0$). The initial surge of seniors crowds out some junior hiring, but because q_{AI} is low, this hiring cut does not severely decrease the pipeline. The senior workforce increases gradually while the junior workforce decreases toward the new steady state. Consequently, the span of control decreases over time (Figure 8a).

If the new learning probability q_{AI} is large (case b and Figure 7b), the transition oscillates. In this case, crowding-out dominates ($S' < 0$). The initial excess of seniors at $t = 2$ also forces the firm to cut junior hiring. But because q is high, this cut destroys a significant portion of the future pipeline, causing a senior shortage at $t = 3$. This shortage forces the firm to hire juniors massively again, creating excess seniors at $t = 4$. Consequently, the span of control oscillates (Figure 8b).

4.5 Discussion and Implications

The previous analysis shows a common pattern across all three types of AI shocks. Whether AI improves productivity (γ^h, γ^l) or learning (q), the short-run impact is always a shift toward a lower span of control. This shift happens immediately for productivity shocks, as firms require fewer juniors, and with one-period lag for learning shocks, as firms accumulate more seniors and consequently require fewer juniors.

This similarity among these three shocks, however, does not imply the same firm structure in the long run. As Section 3 shows, productivity shocks (either γ^h or γ^l) eventually lead the firm back to its original span of control, merely scaling down the total workforce. In contrast, learning shocks permanently result in a lower span.

Furthermore, following the initial shock, the convergence to the new steady state is not always smooth and monotonic. It may involve oscillations driven by the talent pipeline, as the firm alternates between shortage and oversupply of new seniors. Combining previous results, these oscillations happen if and only if, post AI, the crowding-out effect of the pipeline dominates incumbent retention ($S' < 0$):

$$\frac{\gamma^h}{\gamma^l} q(1 - d_l) > 1 - d_h.$$

This inequality identifies three factors that have implications on whether the transition is smooth or oscillating.

First, oscillations are more likely in industries where the productivity gap between seniors and juniors is large, or in industries AI enlarges this gap significantly. A wide gap γ^h/γ^l implies that one senior replaces many juniors. This high substitution rate amplifies the crowding-out effect. Conversely, in industries with a small gap, or AI compresses the gap, firms are more likely to adjust smoothly.[10]

Second, oscillations are more likely in industries where learning is rapid, or where AI accelerates learning significantly. A high learning rate q increases the sensitivity of the senior supply tomorrow to the junior hiring today, amplifying the crowding-out effect. Conversely, in industries where learning is slow even with AI, the pipeline is less sensitive to current hiring, leading to a smooth transition.

Third, oscillations are more likely in industries with high senior attrition but low junior attrition. A high senior attrition (d_h) prevents the firm from relying on incumbents to smooth out pipeline shocks. Meanwhile, a low junior attrition (d_l) ensures that most hires survive to promotion, making future supply highly sensitive to current hiring. This combination characterizes industries where workers join and stay to acquire skills but leave the market shortly after mastering them. In contrast, if senior attrition is low or junior attrition is high, the transition is more likely to be smooth.

[10] See Chen et al. (2025) for evidence of this compression effect of AI.

5 General Equilibrium with Firm-Specific Human Capital

The previous sections assumed that senior workers are perfect substitutes across firms. However, internal experience often creates firm-specific human capital that external hires lack, such as knowledge of workflows, internal protocols, or informal relationships. We now extend the framework to include firm-specific human capital.

5.1 Setup

We distinguish between two types of senior workers. Internal seniors (promoted from within) possess specific capital and have productivity γ^h. External seniors (hired from the outside market) lack this capital and produce only $\gamma^h - \sigma$, where $\sigma > 0$. We assume $\gamma^h - \sigma > \gamma^l$, so that external seniors are still more productive than juniors.

Although productivity differs, the labor market remains competitive. Because seniors lose their specific capital upon switching firms, their outside option is the wage for generic external seniors. Firms therefore pay the market wage w_t^h to all seniors, extracting the rent σ generated by the internal ones.

Let $n_{i,t}^{h,in}$ denote the number of internal seniors, $n_{i,t}^{h,ex}$ the external seniors, and $n_{i,t}^l$ the juniors employed by firm i in period t. The firm maximizes the present discounted value of profits:

$$\max_{\{n_{i,t}^l, n_{i,t}^{h,in}, n_{i,t}^{h,ex}\}_{t=0}^{\infty}} \quad \sum_{t=0}^{\infty} \delta^t \left[Ny - w_t^h(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) - w_t^l n_{i,t}^l \right],$$

subject to the capacity constraint[11]

$$\gamma^h n_{i,t}^{h,in} + (\gamma^h - \sigma) n_{i,t}^{h,ex} + \gamma^l n_{i,t}^l \geq N,$$

and the talent constraint[12]

$$n_{i,t+1}^{h,in} \leq (1 - d_h)(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) + q(1 - d_l) n_{i,t}^l.$$

The talent constraint distinguishes this problem from the benchmark general equilibrium. In the benchmark, periods were independent. Here, hiring a junior in period t generates the future value of a high-productivity senior in period $t + 1$.

[11] From Assumption 1, we have the same logic as that in Lemma 1: value of output is so high that all the tasks should be finished.

[12] We make an implicit assumption here that the external seniors will become internal seniors in the next period.

5.2 Stationary Equilibrium

In a stationary equilibrium with firm-specific human capital, the labor market shows the following properties.

Lemma 4. *With firm-specific human capital, the stationary equilibrium is unique and symmetric. In this equilibrium, every firm hires senior workers exclusively from its own internal labor market* ($n_i^{h,ex} = 0$)*. Furthermore, both the capacity constraint and the talent constraint are binding.*

The intuition for this result comes from the productivity advantage of internal seniors. These workers possess specific human capital that makes them more productive than external hires. Because internal seniors are more productive, firms can capture a rent by promoting from within rather than hiring from the outside market.

The firm also ensures that its capacity and talent constraints bind. In a stationary equilibrium, if the capacity constraint is slack, that means the firm is always paying for idle workers who produce nothing. In such a case, the firm could increase profits by reducing its workforce. If the talent constraint is slack, that means the firm is letting valuable internal talent leave for other firms, which is also suboptimal. As firms are identical and hire senior workers exclusively from within, the binding capacity and talent constraints lead to a unique and symmetric equilibrium.

In this equilibrium, with firm-specific human capital, the wage ratio between seniors and juniors includes not just current productivity, but also the discounted value of the specific capital σ generated in the future.

Lemma 5. *In a stationary symmetric equilibrium, the wage ratio is given by:*

$$\frac{w^h}{w^l} = \frac{(\gamma^h - \sigma) + \delta(1 - d_h)\sigma}{\gamma^l + \delta q(1 - d_l)\sigma} =: \Gamma.$$

Lemma 5 generalizes the wage ratio established in Lemma 2 by incorporating two features. First, the senior wage is tied to the external market. While our earlier general equilibrium benchmark implied this relationship, the inclusion of specific human capital makes it clear that the senior wage must track the productivity of external hires to satisfy the no-arbitrage condition. Second, the ratio incorporates future value into current compensation. By adding the discounted term $\delta(1 - d_h)\sigma$ to the numerator, we account for the future benefit of retaining an external senior (who becomes internal in the future). Correspondingly, the denominator also adds the future value of promotion $\delta q(1 - d_l)\sigma$ for juniors. By including these terms, the ratio captures how today's wages reflect the long-term career path within the firm.

With the wage ratio of labor Γ, we characterize the equilibrium hiring and wage levels which mirror those in the last section (Lemma 3).

Lemma 6. *The stationary symmetric equilibrium is characterized as follows:*

(i) *The equilibrium span of control and the equilibrium employment levels are:*

$$s^* = \frac{d_h}{q(1-d_l)}, \quad n^{l*} = \frac{s^*}{\gamma^l s^* + \gamma^h} N, \quad n^{h,in*} = \frac{1}{\gamma^l s^* + \gamma^h} N, \quad \textit{and} \;\; n^{h,ex*} = 0.$$

(ii) *The equilibrium wages are:*

$$w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\Gamma\delta q(1-d_l)}\right] \quad \textit{and} \quad w^{h*} = \Gamma w^{l*}.$$

5.3 Equilibrium Analysis of AI

We now analyze how AI shocks affect the economy with firm-specific capital. We consider the same three effects: senior productivity (γ^h), junior productivity (γ^l), and learning (q).

Because firms in a stationary symmetric equilibrium hire exclusively from within, the AI shock affects the long-run firm structure ($n^{h,in*}, n^{l*}, s^*$) and wages levels (w^{l*} and w^{h*}) as before. The impact on the wages ratio, however, is different. In the general equilibrium without specific human capital, a learning shock left the wage ratio w^h/w^l constant. Here, faster learning compresses the wage gap.

Corollary 1. *In the steady state, if the learning probability q increases, the ratio between senior wage and junior wage $\frac{w^{h*}}{w^{l*}}$ decreases.*

This compression occurs because the juniors provide not just current production but also future firm-specific capital. A higher learning rate accelerates the creation of this capital, increasing the value of the junior layer relative to the seniors. Consequently, as the juniors become more valuable, the relative wage premium for senior workers declines.

The steady state identifies the destination, but the transition path depends on the firm's hiring strategy over time. The following lemma characterizes the optimal hiring for the firms in a symmetric equilibrium.

Lemma 7. *In a symmetric equilibrium, firms rely exclusively on internal promotion ($n_t^{h,ex} = 0$). Both the capacity and talent constraints bind throughout the transition.*

Lemma 7 shows that, along the transition path after the shock, firms prefer internal hiring. This is because promoted juniors possess specific capital σ, making them more productive than external seniors. Moreover, firms also refrain from hiring more workers than necessary to meet the current production target just for the future value. Because the revenue from any task completed beyond N is zero, the current marginal revenue of an extra worker is zero. Although a junior worker provides future value by filling the talent pipeline, this future value is less than the worker's current wage. As a result, the firm minimizes costs by keeping the capacity constraint and talent constraint binding in every period.

With the firm constrained by its internal pipeline and a fixed output target, the evolution of the senior workforce follows a deterministic path. By substituting the aggregate capacity constraint into the flow equation, we obtain the flow equation for senior workers:

$$n_{t+1}^{h,in} = \left[1 - d_h - \frac{\gamma^h}{\gamma^l} q(1-d_l)\right] n_t^{h,in} + q(1-d_l)\frac{N}{\gamma^l}.$$

This difference equation is identical to the one derived in Section 4. So the dynamics established in Propositions 2, 3, and 4 apply here without modification.

The transition analysis shows that AI shocks can cause the firm structure to oscillate, potentially creating periods where the internal talent pipeline is thin. One might be interested in how these structural disruptions affect the firm's value during the transition. To address this, we derive the firm's value function in the symmetric equilibrium.

Lemma 8. *In the symmetric equilibrium, the value of a firm with $n_t^{h,in}$ internal senior workers is given by:*

$$V(n_t^{h,in}) = \lambda\sigma n_t^{h,in} + \frac{N(y-\lambda)}{1-\delta},$$

where $\lambda = \frac{w^{l*}}{\gamma^l + \delta\sigma q(1-d_l)}$, with w^{l*} being the equilibrium junior wage in Lemma 6.

Lemma 8 decomposes the firm's value into two components, the internal talent value $\lambda\sigma n_t^{h,in}$ and the projected value $\frac{N(y-\lambda)}{1-\delta}$. The projected value represents the discounted stream of rents from the production target N, given the market cost of capacity λ. The talent value represents the premium the firm extracts from its current internal workforce $n_t^{h,in}$, who possess specific capital σ that external hires lack.

Next, we analyze the dynamic impact of AI on firm value throughout the transition.

Proposition 5. *Suppose that productivity (γ^h or γ^l) or the learning probability (q) permanently increases at $t = 1$.*

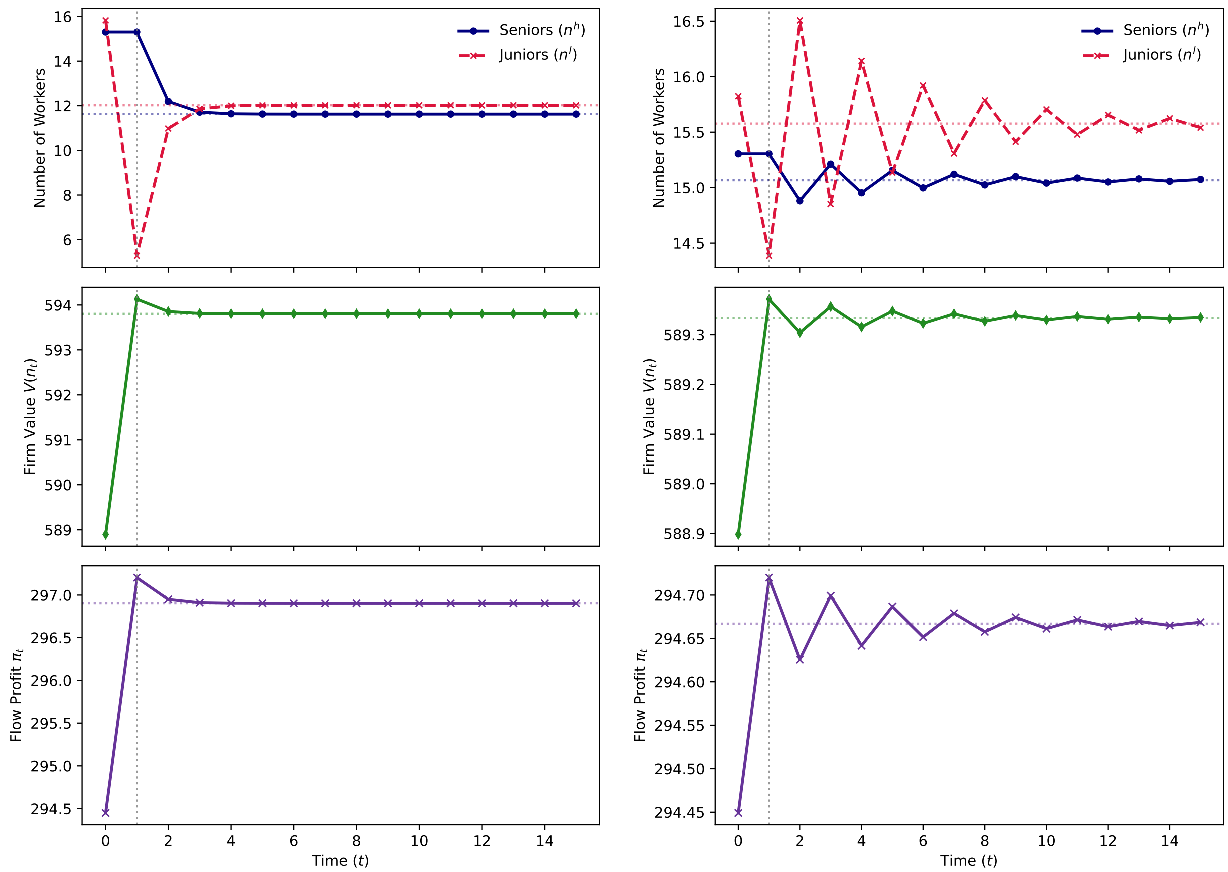
(*i*) *At $t = 1$ the firm's value strictly increases $\left(V(n_1^{h,in}) > V(n_0^{h,in})\right)$.*

(*ii*) *The marginal value of an internal senior, $\lambda\sigma$, falls.*

(*iii*) *During the transition* $(t > 1)$, *the firm's value* $V(n_t^{h,in})$ *comoves with the senior stock* $n_t^{h,in}$.

Proposition 5 shows that the AI shock strictly increases the firm's value at $t = 1$ (Figure 9). Because the market cost of capacity falls forever, the projected value jumps immediately. This gain outweighs the temporary pipeline breakdown, the future shortage of seniors caused by fewer juniors today.

Yet, the transition may entail temporary inefficiency when oscillations happen, or during a temporary pipeline breakdown: when the senior stock dips below the new steady state, the firm captures fewer specific capital rents (σ). This is reflected by the lower flow profit and firm value at $t = 2, 4, \cdots$, compared to the new steady state in Figure 9b.



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure 9: Simulation of the Transition of Discounted Firm Value and Flow Profit. Parameters: $y = 3$, $\delta = 0.5$, $\sigma = 2.45$, $\underline{v} = 0.25$, and the remaining parameters are same as Figure 6. For other shocks, see Figures 10 and 11 in Appendix.

6 Concluding Remarks

Despite its simplicity, our model captures some typical features of ILM, in particular, the need of firms to employ juniors to fill the talent pipeline. Our model is not a typical knowledge hierarchy as in Garicano (2000) and Garicano and Rossi-Hansberg (2015). It is rather a top-down hierarchy in which a CEO implements strategy by allocating tasks to agents in the different hierarchical layers, or, equivalently, delegates work packages that seniors can delegate further to the junior workers. The model also allows for thinking about the different ways in which AI affects workplaces, individual productivity gains on different layers and the speed of learning (here, modeled through the short cut of how likely a junior person acquires the skills needed to become senior). We have abstracted from incentive issues that are certainly an important element of talent management [13] and may need complementary analysis. We hope though that this model is a good start to think about these important issues and would see it as an advantage that it can readily be embedded in a labor market. We are also confident that we can allow for changes in the business model in the industry and are working on it.

References

Arrow, Kenneth J. 1962. "The economic implications of learning by doing." *The review of economic studies*, 29(3): 155–173. 1

Auriol, Emmanuelle, Guido Friebel, and Frauke Von Bieberstein. 2016. "The firm as the locus of social comparisons: Standard promotion practices versus up-or-out." *Journal of Economic Behavior & Organization*, 121: 41–59. 3

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994*a*. "The internal economics of the firm: Evidence from personnel data." *The Quarterly Journal of Economics*, 109(4): 881–919. 1, 3

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994*b*. "The wage policy of a firm." *The Quarterly Journal of Economics*, 109(4): 921–955. 3

Bar-Isaac, Heski, and Clare Leaver. 2021. "Investment in Human Capital Under Endogenous Asymmetric Information." 25

[13] There is a large literature on this, for instance, Waldman (1984); Milgrom and Oster (1987); Carmichael (1988); Fairburn and Malcomson (1994); Prendergast and Topel (1996); Bar-Isaac and Leaver (2021). Recently, many researchers investigate more on the role of manager's incentives (Hoffman and Tadelis, 2021; Friebel, Heinz and Zubanov, 2022; Minni, 2023; Haegele, 2024; Raith and Friebel, 2026)

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. "Generative AI without guardrails can harm learning: Evidence from high school mathematics." *Proceedings of the National Academy of Sciences*, 122(26): e2422633122. 4

Beckmann, Martin. 1978. *Rank in organizations.* Berlin:Springer. 4

Bianchi, Nicola, Giulia Bovini, Jin Li, Matteo Paradisi, and Michael Powell. 2023. "Career spillovers in internal labour markets." *The Review of Economic Studies*, 90(4): 1800–1831. 3, 4

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Stanford Digital Economy Lab Working Paper. 1

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, 140(2): 889–942. 4

Carmichael, H Lorne. 1988. "Incentives in academics: why is there tenure?" *Journal of political Economy*, 96(3): 453–472. 25

Chen, Yvonne Jie, Jie Gong, Jin Li, and Zibo Zhao. 2025. "Better Technology, Worse Motivation: GenAI's Mediocrity Trap." *Available at SSRN 5208163*. 4, 19

Doeringer, Peter B., and Michael J. Piore. 1971. *Internal Labor Markets and Manpower Analysis.* Lexington, Mass:Heath. 1, 3

Fairburn, James A, and James M Malcomson. 1994. "Rewarding performance by promotion to a different job." *European Economic Review*, 38(3-4): 683–690. 25

Friebel, Guido, and Elena Panova. 2008. "Insider Privatization and Careers: A Clinical Study of a Russian Firm in Transition." In *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*. , ed. Stefan Bender, Julia Lane, Kathryn L. Shaw, Fredrik Andersson and Till von Wachter, 253–266. Chicago, IL:University of Chicago Press. 4

Friebel, Guido, Matthias Heinz, and Nikolay Zubanov. 2022. "Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain." *Management Science*, 68(1): 211–229. 25

Fudenberg, Drew, and Luis Rayo. 2019. "Training and effort dynamics in apprenticeship." *American Economic Review*, 109(11): 3780–3812. 4

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of political economy*, 108(5): 874–904. 25

Garicano, Luis, and Esteban Rossi-Hansberg. 2015. "Knowledge-based hierarchies: Using organizations to understand the economy." *Annual Review of Economics*, 7(1): 1–30. 25

Garicano, Luis, and Luis Rayo. 2017. "Relational knowledge transfers." *American Economic Review*, 107(9): 2695–2730. 4

Garicano, Luis, and Luis Rayo. 2025. "Training in the Age of AI: A Theory of Apprenticeship Viability." Centre for Economic Policy Research Discussion Paper DP20634. 4

Gibbons, Robert. 1997. "Incentives and Careers in Organizations." In *Advances in Econometrics: Theory and Applications*. Vol. 2, , ed. David M. Kreps and Kenneth J. Wallis, Chapter 1, 1–37. Cambridge:Cambridge University Press. 3

Gibbons, Robert, and Michael Waldman. 1999*a*. "Careers in organizations: Theory and evidence." In *Handbook of Labor Economics*. Vol. 3, 2373–2437. Elsevier. 1, 3

Gibbons, Robert, and Michael Waldman. 1999*b*. "A theory of wage and promotion dynamics inside firms." *The Quarterly Journal of Economics*, 114(4): 1321–1358. 3, 42

Gibbons, Robert, and Michael Waldman. 2004. "Task-specific human capital." *American Economic Review*, 94(2): 203–207. 1

Haegele, Ingrid. 2024. "Talent hoarding in organizations." *arXiv preprint arXiv:2206.15098*. 25

Hoffman, Mitchell, and Steven Tadelis. 2021. "People management skills, employee attrition, and manager rewards: An empirical analysis." *Journal of political economy*, 129(1): 243–285. 25

Huitfeldt, Ingrid, Andreas R Kostøl, Jan Nimczik, and Andrea Weber. 2023. "Internal labor markets: A worker flow approach." *Journal of Econometrics*, 233(2): 661–688. 4

Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2024. "The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market." *Organization Science*, 35(6): 1977–1989. 1

Ide, Enrique, and Eduard Talamàs. 2025. "Artificial Intelligence in the Knowledge Economy." *Journal of Political Economy*, 133(12): 3762–3800. 4

Ke, Rongzhu, Jin Li, and Michael Powell. 2018. "Managing careers in organizations." *Journal of Labor Economics*, 36(1): 197–252. 3, 4

Kräkel, Matthias, and Anja Schöttner. 2012. "Internal labor markets and worker rents." *Journal of Economic Behavior & Organization*, 84(2): 491–509. 3

Lazear, Edward P., and Paul Oyer. 2013. "Personnel Economics." In *Handbook of Organizational Economics*. , ed. Robert Gibbons and John Roberts, Chapter 12, 479–519. Princeton, NJ:Princeton University Press. 3

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89(5): 841–864. 3

Lichtinger, Guy, and Seyed Mahdi Hosseini Maasoum. 2025. "Generative AI as seniority-biased technological change: Evidence from US resume and job posting data." *Available at SSRN*. 1

MacLeod, W Bentley, and James M Malcomson. 1988. "Reputation and hierarchy in dynamic models of employment." *Journal of Political Economy*, 96(4): 832–854. 3

Malcomson, James M. 1984. "Work incentives, hierarchy, and internal labor markets." *Journal of political economy*, 92(3): 486–507. 3

Milgrom, Paul, and Sharon Oster. 1987. "Job discrimination, market forces, and the invisibility hypothesis." *The Quarterly Journal of Economics*, 102(3): 453–476. 25

Minni, Virginia. 2023. "Making the invisible hand visible: Managers and the allocation of workers to jobs." 25

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science*, 381(6654): 187–192. 4

Osterman, Paul. 2024. "Reconsidering occupational internal labor markets: Incidence and consequences." *ILR Review*, 77(3): 366–395. 4

Pastorino, Elena. 2024. "Careers in firms: The role of learning about ability and human capital acquisition." *Journal of Political Economy*, 132(6): 1994–2073. 3

Prendergast, Canice. 1993. "The role of promotion in inducing specific human capital acquisition." *The Quarterly Journal of Economics*, 108(2): 523–534. 3

Prendergast, Canice, and Robert H Topel. 1996. "Favoritism in organizations." *Journal of Political Economy*, 104(5): 958–978. 25

Priest, Dan. 2025. "In the age of AI: Speed matters more, scale matters less, innovation matters most." *PwC Tech Effect*, Accessed: January 14, 2026. 1

Raith, Michael, and Guido Friebel. 2026. "Talent Management: The Role of Bosses." *Journal of Labor Economics*. Forthcoming. 25

Rosenbaum, James. 1984. *Career mobility in a corporate hierarchy.* New York:Academic Press. 4

Rosen, Sherwin. 1986. "Prizes and incentives in elimination tournaments." *American Economic Review*, 76(4): 701–715. 3

Sharwood, Simon. 2025. "AWS CEO says using AI to replace junior staff is 'Dumbest thing I've ever heard'." 1

Simon, Herbert A. 1951. "The analysis of promotional opportunities." *Personnel*, 27(4): 282–285. 4

Stewman, Shelby, and Suresh Konda. 1983. "Careers and organizational labor markets: Demographic models of organizational behavior." *American Journal of Sociology*, 88(4): 637–685. 4

Waldman, Michael. 1984. "Job assignments, signalling, and efficiency." *The Rand journal of economics*, 15(2): 255–267. 25

Waldman, Michael. 2003. "Ex ante versus ex post optimal promotion rules: The case of internal promotion." *Economic Inquiry*, 41(1): 27–41. 3

Waldman, Michael. 2013. "Theory and evidence in internal labor markets." In *The handbook of organizational economics*. , ed. Robert Gibbons and John Roberts, Chapter 13, 520–571. Princeton University Press. 1, 3

White, Harrison C. 1970. *Chains of opportunity: System models of mobility in organizations.* Harvard University Press. 4

Zabojnik, Jan, and Dan Bernhardt. 2001. "Corporate tournaments, human capital acquisition, and the firm Size—Wage relation." *The Review of Economic Studies*, 68(3): 693–716. 3

A Proof

Proof of Lemma 1. We solve the simplified optimization problem using the Lagrangian function:

$$
\begin{aligned}
\mathcal{L} =& (D^l_{i,t} + D^h_{i,t})y - w^h_t n^h_{i,t} - w^l_t n^l_{i,t} \\
&+ \lambda_{1,t}(\gamma^l n^l_{i,t} - D^l_{i,t}) \\
&+ \lambda_{2,t}(\gamma^h n^h_{i,t} - D^h_{i,t}) \\
&+ \lambda_{3,t}(N - D^l_{i,t} - D^h_{i,t}),
\end{aligned}
$$

where $\lambda_{1,t}$, $\lambda_{2,t}$ and $\lambda_{3,t}$ are Lagrangian multipliers for the capacity constraints (CC). The first order conditions on the four control variables are

$$
\begin{aligned}
n^l_{i,t}:& \quad -w^l_t + \lambda_{1,t}\gamma^l = 0; \\
n^h_{i,t}:& \quad -w^h_t + \lambda_{2,t}\gamma^h = 0; \\
D^l_{i,t}:& \quad y - \lambda_{1,t} - \lambda_{3,t} = 0; \\
D^h_{i,t}:& \quad y - \lambda_{2,t} - \lambda_{3,t} = 0.
\end{aligned}
$$

This shows that $\lambda_{1,t} = w^l_t/\gamma^l > 0$ and $\lambda_{2,t} = w^h_t/\gamma^h > 0$, meaning that $\gamma^l n^l_{i,t} = D^l_{i,t}$ and $\gamma^h n^h_{i,t} = D^h_{i,t}$. Moreover, $\lambda_{3,t} = y - \lambda_{1,t} = y - w^l_t/\gamma^l$. From ($VF^l$), we have

$$
w^l_t = v^l_t - \delta(1-d_l)[qv^h_{t+1} + (1-q)v^l_{t+1}] + \delta d_l \underline{v}.
$$

Free entry condition makes sure that $v^l_t = v^l_{t_1} = \underline{v}$. Since $v^h_{t+1} \geq \underline{v}$, we must have $w^l_t \leq (1-\delta)\underline{v}$. Together with Assumption 1, we have $\lambda_{3,t} > 0$, meaning that $D^l_{i,t} + D^h_{i,t} = N$. ■

Proof of Lemma 2. From the proof of Lemma 1, we have $\lambda_{1,t} = y - \lambda_{3,t} = \lambda_{2,t}$. This means $w^l_t/\gamma^l = w^h_t/\gamma^h$. ■

Proof of Lemma 3. Part (i): In a symmetric equilibrium, the flow condition for a representative firm is given by

$$
n^h_{i,t+1} = (1-d_h)n^h_{i,t} + q(1-d_l)n^l_{i,t}.
$$

Together with the steady state constraints (SS), we have

$$
s^* = \frac{n^{l*}}{n^{h*}} = \frac{d_h}{q(1-d_l)}.
$$

Lemma 1 shows that all the capacity constraints (CC) bind. This means $n^l_{i,t} = (N - \gamma^h n^h_{i,t})/\gamma^l$.

Hence, we have

$$n^{h*} = (1-d_h)n^{h*} + \frac{q}{\gamma^l}(1-d_l)(N - \gamma^h n^h_{i,t}),$$

or $n^{h*} = \frac{q(1-d_l)}{\gamma^l d_h + \gamma^h q(1-d_l)} N = \frac{1}{\gamma^l s^* + \gamma^h} N$. Hence, $n^{l*} = (N - \gamma^h n^{h*})/\gamma^l = \frac{d_h}{\gamma^l d_h + \gamma^h q(1-d_l)} = \frac{s^*}{\gamma^l s^* + \gamma^h} N$.

Part (ii): From Lemma 2, we have $w^l_t/\gamma^l = w^h_t/\gamma^h$. Therefore, in a steady state, we must have

$$w^{h*} = \frac{\gamma^h}{\gamma^l} w^{l*}.$$

With the steady state and free entry, we have $v^l_t = V^l_{t+1} = \underline{v}$ and $v^h_t = v^h_{t+1}$. Together with the value functions (VF^l) and (VF^h), we have

$$\underline{v} = w^{l*} + \delta(1-d_l)[q v^h_t + (1-q)\underline{v}] + \delta d_l \underline{v},$$
$$v^h_t = \frac{\gamma^h}{\gamma^l} w^{l*} + \delta(1-d_h)v^h_t + \delta d_h \underline{v}.$$

Hence, we can get $w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)}\right]$. ■

Proof of Proposition 1. These results are directly from Lemma 3.

Part (i):

$$\frac{\partial n^{h*}}{\partial \gamma^h} < 0, \quad \frac{\partial n^{l*}}{\partial \gamma^h} < 0, \quad \frac{\partial s^*}{\partial \gamma^h} = 0, \quad \frac{\partial w^{l*}}{\partial \gamma^h} < 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial \gamma^h} > 0.$$

Part (ii):

$$\frac{\partial n^{h*}}{\partial \gamma^l} < 0, \quad \frac{\partial n^{l*}}{\partial \gamma^l} < 0, \quad \frac{\partial s^*}{\partial \gamma^l} = 0, \quad \frac{\partial w^{l*}}{\partial \gamma^l} > 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial \gamma^l} < 0.$$

Part (iii):

$$\frac{\partial n^{h*}}{\partial q} > 0, \quad \frac{\partial n^{l*}}{\partial q} < 0, \quad \frac{\partial s^*}{\partial q} < 0, \quad \frac{\partial w^{l*}}{\partial q} < 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial q} < 0.$$

From Lemma 2, $w^{h*}/w^{l*} = \gamma^h/\gamma^l$, meaning that a change in q has no effect on the wage ratio. ■

Proof of Proposition 2. Focusing on the symmetric equilibrium, we have the following flow condition for a representative firm:

$$n^h_{t+1} = (1-d_h)n^h_t + q(1-d_l)n^l_t. \qquad (FC^{ind})$$

We analyze the transition dynamics using the individual firm flow condition (FC^{ind}) and the binding capacity constraints established in Lemma 1.

Part (i) Impact at $t = 1$. At the beginning of $t = 1$, the number of seniors is given by the flow from the initial steady state. Thus, the senior workforce at $t = 1$ remains at the pre-shock level.

Firm optimality requires capacity constraints to bind. With the new productivity γ^h_{AI}, the demand for junior workers at $t = 1$ is:

$$n^l_1 = \frac{N - \gamma^h_{AI} n^h_1}{\gamma^l} = \frac{N - \gamma^h_{AI} n^{h*}_0}{\gamma^l}.$$

The assumption $\gamma^h_{AI} < N/n^{h*}_0$ implies $N - \gamma^h_{AI} n^{h*}_0 > 0$, ensuring $n^l_1 > 0$. Since $\gamma^h_{AI} > \gamma^h_0$, it follows that $n^l_1 < n^{l*}_0$. The span of control is $s_1 = n^l_1/n^h_1 = n^l_1/n^{h*}_0$. Since $n^l_1 < n^{l*}_0$, we have $s_1 < s^*_0$.

Part (ii) Impact for $t \geq 2$. Substitute $n^l_t = (N - \gamma^h_{AI} n^h_t)/\gamma^l$ into the flow equation (FC^{ind}) to obtain the linear first-order difference equation (S):

$$n^h_{t+1} = \underbrace{\left[(1 - d_h) - \frac{q(1 - d_l)\gamma^h_{AI}}{\gamma^l}\right]}_{S'} n^h_t + \frac{q(1 - d_l)N}{\gamma^l}.$$

The general solution is $n^h_t - n^{h*}_{AI} = (S')^{t-1}(n^h_1 - n^{h*}_{AI})$. Stability requires $|S'| < 1$, which we assume holds.

We must verify that firms employ juniors in every period during the transition. From the capacity constraint, $n^l_t > 0$ if and only if $n^h_t < N/\gamma^h_{AI}$. From Proposition 1, an increase in γ^h reduces the steady-state senior stock, so $n^{h*}_{AI} < n^{h*}_0 = n^h_1$. By assumption, we know $n^h_1 < N/\gamma^h_{AI}$. Because $|S'| < 1$, the distance from the steady state shrinks over time: $|n^h_t - n^{h*}_{AI}| \leq |n^h_1 - n^{h*}_{AI}|$. This implies that n^h_t is bounded by the initial value:

$$n^h_t \leq \max(n^h_1, n^{h*}_{AI}) = n^h_1 < \frac{N}{\gamma^h_{AI}}.$$

Therefore, $n^l_t > 0$ for all t.

Next we show the transition path. The transition path is determined by the sign of S':

Case (a): If $\gamma^h_{AI} \leq \frac{1-d_h}{q(1-d_l)}\gamma^l$, then $S' \in [0, 1)$. The deviation $(n^h_t - n^{h*}_{AI})$ maintains the same sign. Since $n^h_1 > n^{h*}_{AI}$, the sequence $\{n^h_t\}$ decreases monotonically toward n^{h*}_{AI}. Since n^l_t is linearly decreasing in n^h_t, n^l_t increases monotonically. The span of control $s_t = n^l_t/n^h_t$ therefore increases monotonically.

Case (b): If $\gamma^h_{AI} > \frac{1-d_h}{q(1-d_l)}\gamma^l$, then $S' \in (-1, 0)$. The deviation alternates sign: sign$(n^h_{t+1} -$

$n_{AI}^{h*}) = -\text{sign}(n_t^h - n_{AI}^{h*})$. Thus, n_t^h oscillates around the new steady state. Consequently, n_t^l and the span s_t fluctuate. ∎

The proofs for Propositions 3 and 4 follow the same logic and are omitted. The only notable difference is that the condition $n_1^l > 0$ is guaranteed without additional assumptions. For the junior productivity shock, the senior stock at $t = 1$ is unchanged ($n_1^h = n_0^{h*}$), implying aggregate senior output remains strictly below N. For the learning shock, neither stocks nor productivities change at $t = 1$, so hiring remains at the initial steady state level ($n_1^l = n_0^{l*} > 0$).

Proof of Lemma 4. Consider a firm maximizes the present discounted value of profits subject to the capacity constraint and the talent flow constraint. Let λ_t and μ_t denote the Lagrange multipliers for the capacity and talent constraints, respectively. The Lagrangian is:

$$\mathcal{L} = \sum_{t=0}^{\infty} \delta^t \left\{ \begin{aligned} & Ny - w_t^h(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) - w_t^l n_{i,t}^l \\ & + \lambda_{i,t}[\gamma^h n_{i,t}^{h,in} + (\gamma^h - \sigma) n_{i,t}^{h,ex} + \gamma^l n_{i,t}^l - N] \\ & + \mu_{i,t}[(1-d_h)(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) + q(1-d_l)n_{i,t}^l - n_{i,t+1}^{h,in}] \end{aligned} \right\}.$$

In a stationary equilibrium, the wages and multipliers are constant ($w_t^h = w^h$, $w_t^l = w^l$, $\lambda_{i,t} = \lambda_i$, $\mu_{i,t} = \mu_i$).

i. We first prove that $n_{it}^{h,ex} = 0$ so that $n_{it}^l > 0$ and $n_{it}^{h,in} > 0$. Suppose, by contradiction, that there exists a stationary equilibrium where the aggregate mass of external seniors hired is strictly positive ($\int n_i^{h,ex} di > 0$).

(i) There exists a firm i that hires external seniors ($n_i^{h,ex} > 0$). For this firm, the talent constraint must be binding. Otherwise, that means the firm lets some internal seniors go away while hiring externally. Then firm i can increase the internal hiring and reduces external hiring in a way to keep the capacity constant. This would strictly lower the wage bill.

Now we know that, for this firm i, $n_i^{h,ex} > 0$ and $n_i^{h,in} > 0$. Consequently, the first-order conditions with respect to $n_i^{h,in}$ and $n_i^{h,ex}$ (which are strictly positive) are:

$$\mu_i = \delta[-w^h + \lambda_i \gamma^h + \mu_i(1-d_h)]. \tag{1}$$

$$w^h = \lambda_i(\gamma^h - \sigma) + \mu_i(1-d_h), \tag{2}$$

Substituting (2) into (1) yields $\mu_i = \delta\lambda_i\sigma$, which captures the present value of the specific human capital premium σ. We substitute $\mu_i = \delta\lambda_i\sigma$ back into the first-order condition for

$n_{i,t}^{h,ex}$ and the condition for $n_{i,t}^{l}$ (which is nonnegative):

$$w^h = \lambda_i[\gamma^h - \sigma + \delta\sigma(1-d_h)],$$
$$w^l \geq \lambda_i[\gamma^l + \delta\sigma q(1-d_l)].$$

Taking the ratio of wages leads to the following upper bound (we show later wages are strictly positive):

$$\frac{w^h}{w^l} \leq \frac{\gamma^h - \sigma[1-\delta(1-d_h)]}{\gamma^l + \delta\sigma q(1-d_l)}. \tag{3}$$

Since $\sigma > 0$, the numerator is strictly less than γ^h, and the denominator is strictly greater than γ^l. Thus, for any firm i hiring external seniors, $\frac{w^h}{w^l} < \frac{\gamma^h}{\gamma^l}$.

(ii) On the other hand, for external seniors to be hired, there must be a firm j releasing their internal talent. That is, firm j's talent constraint is slack. This implies that $\mu_j = 0$. First notice that firm j does not hire externally ($n_j^{h,ex} = 0$). If firm j hires externally, then it can always reduce the external hiring by ε and increase internal hiring by $\frac{\gamma^h - \sigma}{\gamma^h}\varepsilon$, as the talent constraint is slack. This keeps the capacity constant while strictly decreasing the wage bill.

Second, notice that firm j employs juniors ($n_j^l > 0$). If $n_j^l = 0$ and $n_j^{h,ex} = 0$, the talent constraint implies $n_j^{h,in} < (1-d_h)n_j^{h,in}$, which is impossible. Moreover, the capacity constraint binds. If not, the firm could reduce costs by firing juniors and strictly raising the profit. This implies that $\lambda_j > 0$.

Because for firm j, $n_j^l > 0$ and $n_j^{h,ex} = 0$, the first-order conditions for firm j with respect to n_j^l and $n_j^{h,ex}$ are

$$w^l = \lambda_j \gamma^l,$$
$$w^h \geq \lambda_j \gamma^h.$$

Therefore, wages are strictly positive. Taking the ratio yields:

$$\frac{w^h}{w^l} \geq \frac{\gamma^h}{\gamma^l}. \tag{4}$$

Comparing (3) and (4), we have $\frac{w^h}{w^l} < \frac{\gamma^h}{\gamma^l} \leq \frac{w^h}{w^l}$, a contradiction. Thus, no firms hire external seniors in stationary equilibrium. Every firm hires juniors and hires seniors only from the ILM ($n_{it}^l > 0$ and $n_{it}^{h,in} > 0$).

ii. Next we prove that the capacity and talent constraints are binding. If the capacity constraint is slack, then each firm can reduce $n_{it}^{h,in}$ and n_{it}^l in proportion to satisfy the flow constraint while strictly reduces the wage bill. Second, suppose the talent constraint is slack for a positive mass of firms ($n_i^{h,in} < (1-d_h)n_i^{h,in} + q(1-d_l)n_i^l$). These firms release talent into the external market, implying the aggregate supply of external seniors is strictly positive. However, we have shown that the demand for external seniors is zero for all firms. This violates the market clearing condition. Thus, the talent constraint must bind: $d_h n_i^{h,in} = q(1-d_l)n_i^l$. Because both capacity and talent constraints are binding, they uniquely pin down $n_t^{h,in}$ and n_t^l in the stationary equilibrium and every firm is symmetric. ■

Proof of Lemma 5. From Lemma 4, we know $n_t^l > 0$, $n_t^{h,in} > 0$, and $n_t^{h,ex} = 0$. Therefore, the first-order condition for junior workers n_t^l is:

$$\frac{\partial \mathcal{L}}{\partial n_t^l} = -w^l + \lambda_t \gamma^l + \mu_t q(1-d_l) = 0.$$

For internal seniors $n_{t+1}^{h,in}$, it is:

$$\frac{\partial \mathcal{L}}{\partial n_{t+1}^{h,in}} = -\mu_t + \delta\left[-w^h + \lambda_{t+1}\gamma^h + \mu_{t+1}(1-d_h)\right] = 0.$$

For external seniors $n_t^{h,ex}$, it is:

$$\frac{\partial \mathcal{L}}{\partial n_t^{h,ex}} = -w^h + \lambda_t(\gamma^h - \sigma) + \mu_t(1-d_h) = 0.$$

The condition holds with equality because of competitive equilibrium.

In the stationary symmetric equilibrium, multipliers are constant ($\lambda_t = \lambda, \mu_t = \mu$). Rearranging these equations yields:

$$\begin{aligned} w^h &= \lambda(\gamma^h - \sigma) + (\delta\lambda\sigma)(1-d_h) = \lambda[\gamma^h - \sigma + \delta\sigma(1-d_h)], \\ w^l &= \lambda\gamma^l + (\delta\lambda\sigma)q(1-d_l) = \lambda[\gamma^l + \delta\sigma q(1-d_l)]. \end{aligned}$$

The multiplier λ is positive because the capacity constraint binds. Dividing w^h by w^l eliminates λ and yields the result Γ. ■

Proof of Lemma 6. Part (i) is followed directly from the two binding constraints derived in Lemma 4. Part (ii) follows from Lemma 5, value functions (VF^l) and (VF^h), and the steady state conditions. ■

Proof of Lemma 7. We first prove that, if there is a symmetric equilibrium involving external hiring, we can always construct a symmetric equilibrium that achieves the same profit for the firm without external hiring but strictly increases firm's capacity.

Suppose that firms hire external seniors during the transition ($n_t^{h,ex} > 0$). In a symmetric equilibrium, the supply of external seniors corresponds to the seniors that firms fail to retain. Thus, a firm hiring $n_t^{h,ex}$ external seniors could explicitly choose to retain these workers instead.

Consider this alternative strategy: the firm sets $\tilde{n}_t^{h,in} = n_t^{h,in} + n_t^{h,ex}$ and $\tilde{n}_t^{h,ex} = 0$. This reallocation maintains the total senior workforce and the wage bill $w_t^h(n_t^{h,in} + n_t^{h,ex})$. Furthermore, because the total number of seniors remains unchanged, the firm carries the exact same talent pool into the next period, satisfying all dynamic constraints. However, because internal seniors possess firm-specific capital while external seniors do not, total capacity increases. Therefore, it is without loss of generality to focus on the symmetric equilibrium with internal hiring only. We next prove that the slack capacity constraint implies that the original plan is suboptimal if the firm's optimal strategy requires the capacity constraint to bind.

To show that the capacity and talent constraints bind, we analyze the firm's dynamic optimization problem. The firm maximizes discounted profits subject to both constraints. Although firms hire only internal seniors in equilibrium, the availability of external hires determines the market wage w_t^h. Let λ_t be the Lagrange multiplier for the capacity constraint and μ_t be the multiplier for the talent constraint. The Lagrangian is:

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \delta^t \Bigg\{ & Ny - w_t^h n_t^{h,ex} - w_t^h n_t^{h,in} - w_t^l n_t^l \\ & + \lambda_t \left[(\gamma^h - \sigma) n_t^{h,ex} + \gamma^h n_t^{h,in} + \gamma^l n_t^l - N \right] \\ & + \mu_t \left[(1 - d_h)(n_t^{h,in} + n_t^{h,ex}) + q(1 - d_l) n_t^l - n_{t+1}^{h,in} \right] \Bigg\}.\end{aligned}$$

The first-order conditions for $n_t^{h,ex}$, n_t^l, and $n_{t+1}^{h,in}$ characterize the firm's optimal hiring:

$$\begin{aligned}-w_t^h + \lambda_t(\gamma^h - \sigma) + \mu_t(1 - d_h) = 0, \\ -w_t^l + \lambda_t \gamma^l + \mu_t q(1 - d_l) = 0, \\ \delta \left[-w_{t+1}^h + \lambda_{t+1}\gamma^h + \mu_{t+1}(1 - d_h) \right] - \mu_t = 0.\end{aligned}$$

These three conditions imply $\mu_t = \delta\lambda\sigma$ and

$$\begin{aligned}w_t^h &= \lambda_t(\gamma^h - \sigma) + \delta\sigma(1 - d_h)\lambda_{t+1}, \\ w_t^l &= \lambda_t \gamma^l + \delta\sigma q(1 - d_l)\lambda_{t+1}.\end{aligned}$$

That is, in a dynamic symmetric equilibrium, the no-arbitrage condition becomes $\frac{w_t^h}{w_t^l} = \frac{\lambda_t(\gamma^h-\sigma)+\delta\sigma(1-d_h)\lambda_{t+1}}{\lambda_t\gamma^l+\delta\sigma q(1-d_l)\lambda_{t+1}}$. In the linear economy with competitive labor markets, shadow prices adjust instantly to the new steady-state levels following a permanent shock: $\lambda_t = \lambda$. As a result, $\frac{w^h}{w^l} = \frac{(\gamma^h-\sigma)+\delta(1-d_h)\sigma}{\gamma^l+\delta q(1-d_l)\sigma}$.

Finally, we need to prove λ (and hence μ) is positive. This is done by solving the wages out from the no-arbitrage condition, workers' value equations, and the free-entry condition as before. The expressions of the wages are the same as those in Lemma 6 except the γ^h, γ^l, and q changes to the new values after the shock. Because market wages are strictly positive, the shadow price λ (and consequently μ) must be strictly positive. By the Kuhn-Tucker conditions, $\lambda > 0$ and $\mu > 0$ imply that both constraints bind at the optimum. This concludes the proof. ■

Proof of Lemma 8. We conjecture a linear value function $V(n^{h,in}) = A + Bn^{h,in}$. Substituting this into the Bellman equation $V(n_t^{h,in}) = Ny - w^h n_t^{h,in} - w^l n_t^l + \delta V(n_{t+1}^{h,in})$ and using the wage expressions from Lemma 5, we match coefficients.

First, the binding capacity constraint implies $n_t^l = (N - \gamma^h n_t^{h,in})/\gamma^l$. The transition law is $n_{t+1}^{h,in} = S' n_t^{h,in} + C$, where S' and C are constants defined by the binding flow constraint. Matching the slope coefficients yields $B = \lambda\sigma$. Matching the constant terms yields $A = \frac{N(y-\lambda)}{1-\delta}$.

The expression λ follows from the junior wage equation derived in the proof of Lemma 5, $w^l = \lambda[\gamma^l + \delta\sigma q(1-d_l)]$. ■

Proof of Proposition 5. For parts (i) and (ii), differentiating λ with respect to the parameters shows that $\frac{\partial\lambda}{\partial\gamma^h} < 0$, $\frac{\partial\lambda}{\partial\gamma^l} < 0$, and $\frac{\partial\lambda}{\partial q} < 0$. This implies the shadow cost of capacity falls with AI improvements.

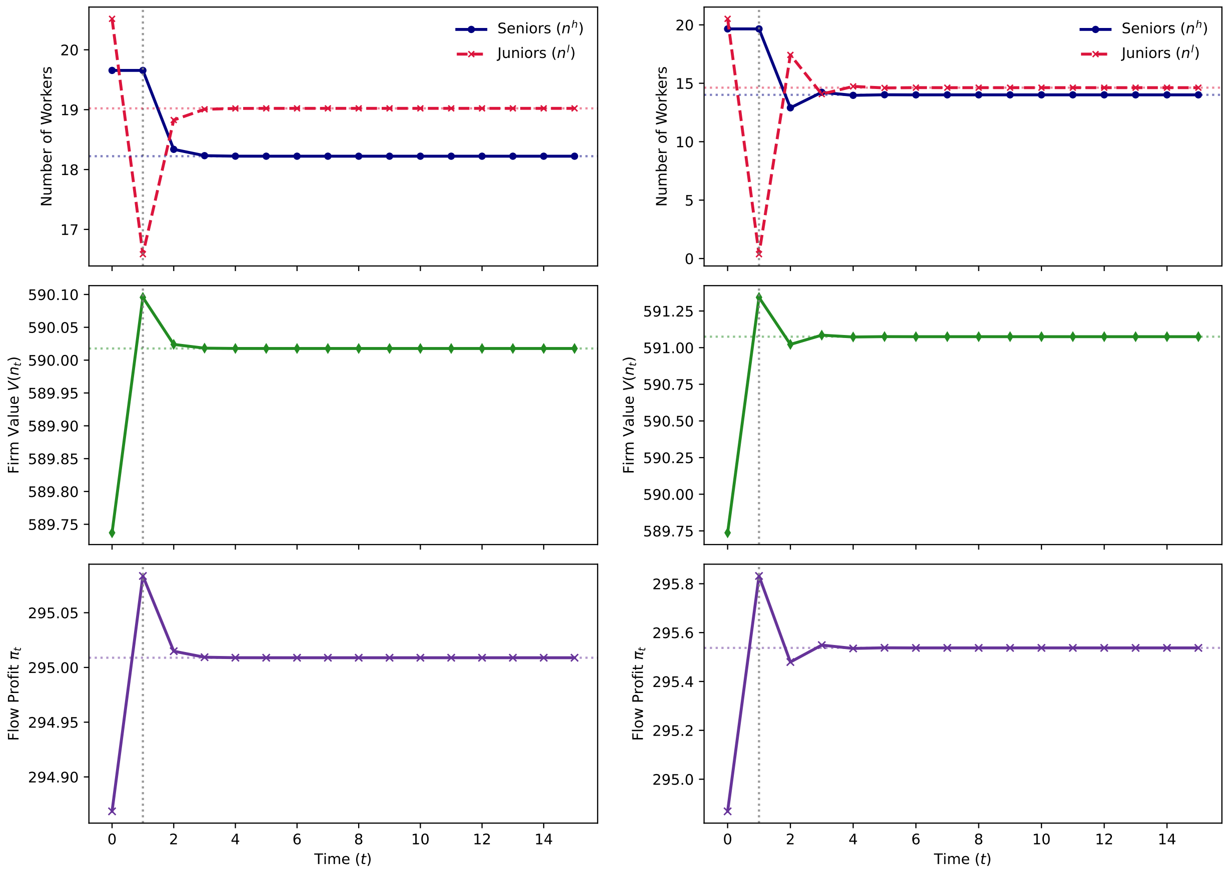
Using the value function $V(n_t^{h,in}) = \lambda\sigma n_t^{h,in} + \frac{N(y-\lambda)}{1-\delta}$, we differentiate with respect to λ:

$$\frac{\partial V}{\partial\lambda} = \sigma n_t^{h,in} - \frac{N}{1-\delta}.$$

Since the firm's total capacity is N and internal seniors produce $\gamma^h > \sigma$, the total specific capital rent cannot exceed the total capacity value, so $\sigma n_t^{h,in} < N < \frac{N}{1-\delta}$. Thus, $\frac{\partial V}{\partial\lambda} < 0$. Because AI lowers λ, and lower λ increases V, the firm value increases instantly. Simultaneously, the marginal value of an incumbent is $\lambda\sigma$, which strictly falls as λ falls.

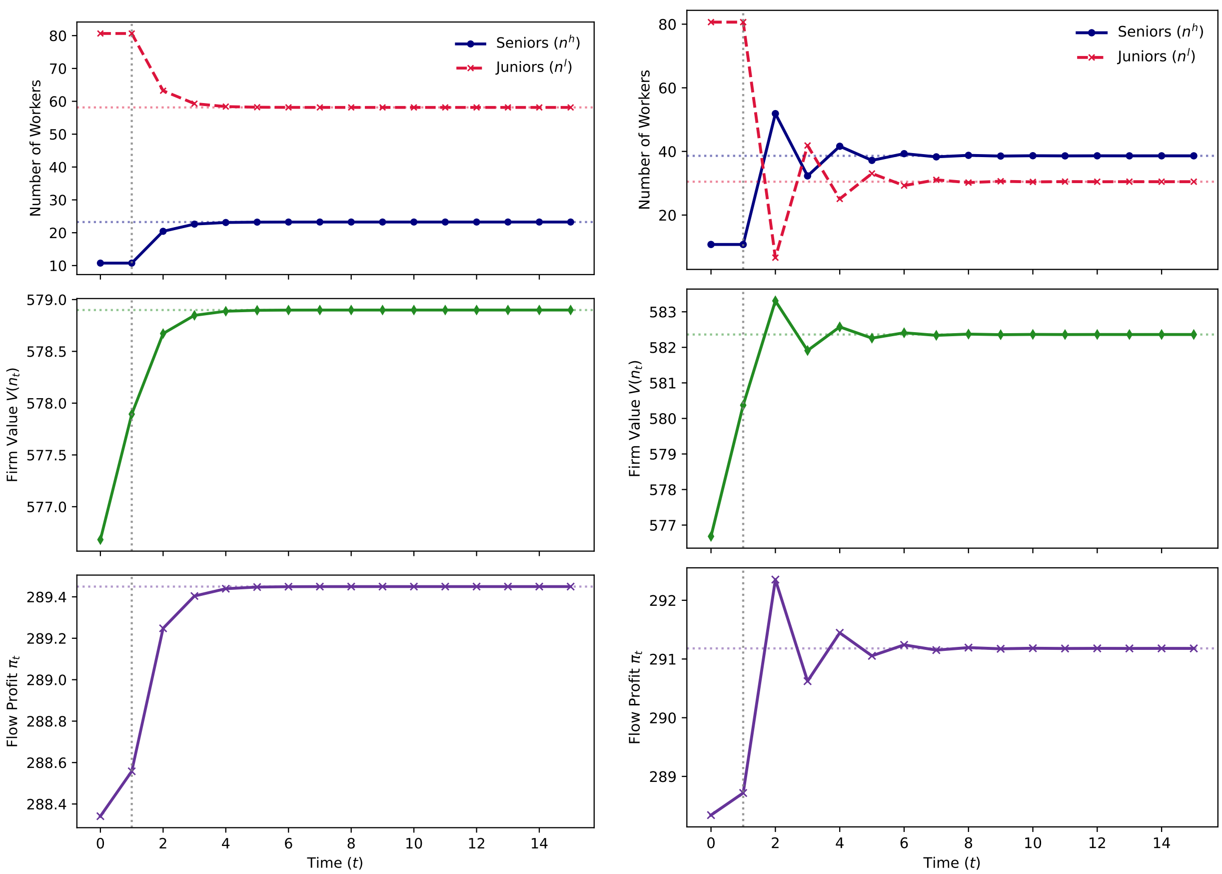
For part (iii), from Lemma 8, $V(n_t^{h,in})$ is linear and increasing in $n_t^{h,in}$ (since $\lambda\sigma > 0$). From the transition analysis, we know $n_t^{h,in}$ can oscillate during the transition. Therefore, $V(n_t)$ can also oscillate. Specifically, if $n_t^{h,in} < n^{h,in*}$, then $V(n_t^{h,in}) < V(n^{h,in*})$. ■

B Figures



(**a**) Smooth Transition. (**b**) Oscillation.

Figure 10: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\sigma = 0.99$, $\underline{v} = 0.25$, and the remaining parameters are same as Figure 4.



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure 11: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\sigma = 0.7$, $\underline{v} = 0.25$, and the remaining parameters are same as Figure 8.

C Extensions

C.1 Extreme AI Shocks and Senior Oversupply

Our previous analysis assumed that the AI productivity shock to seniors was not extreme. In that case, even with more productive seniors, firms still needed junior workers to meet their production targets. We now consider an extreme AI shock where senior productivity γ_h^{AI} rises so much that the existing senior stock can produce more than the market requires ($N_0^{h*} > MN/\gamma_h^{AI}$).

In this case, there exists an oversupply of senior workers. We assume that unemployed workers receive a flow unemployment benefit $b \in (0, (1-\delta)\bar{v})$.[14] The following proposition characterizes the transition when AI creates such an oversupply of seniors.

Proposition C.1. *Consider an AI shock that creates an initial oversupply of senior workers. The transition to the new steady state occurs in three phases:*

(*i*) *If the expected duration of oversupply exceeds a threshold T_{max}, a mass of seniors exits the market immediately at $t = 1$.*

(*ii*) *For a finite number of periods, the senior wage remains at b, and junior hiring is frozen ($N_t^l = 0$). The supply of seniors exceeds demand, causing involuntary unemployment.*

(*iii*) *Once attrition eliminates the senior oversupply, unemployment disappears, the senior wage jumps to its new steady-state level, and firms resume hiring juniors. The subsequent transition follows Proposition 2.*

The logic behind Proposition C.1 follows the seniors' decision to wait or exit. When the stock of senior workers exceeds the firm's demand, competition among seniors drives their wage down to the floor set by the unemployment benefit, b. Facing this low wage, seniors must weigh the cost of waiting against the benefit of future employment. Waiting yields only b today but offers the continuation value of returning to a high-wage job once the oversupply clears.

If the initial stock of seniors is too large, however, the oversupply takes a long time to clear. A long wait reduces the present value of staying in the market below the outside option $\bar{v}$. Consequently, a mass of seniors exits immediately at $t = 1$, until the remaining stock is small enough that the expected wait time justifies staying.

[14] The unemployment benefit b serves as a floor for the market wage. It represents, for example, the flow utility of leisure or home production while searching for a job. We assume $b < (1-\delta)\bar{v}$ to ensure that unemployment is less desirable than the outside option of leaving the market entirely.

For those who stay, the market enters a period where firms freeze junior hiring. Because seniors are now highly productive (γ_h is high) but cost only the minimum wage (b), firms fill all tasks with seniors and stop hiring new entrants entirely ($N_t^l = 0$).

This hiring freeze persists until natural attrition reduces the senior stock. In every period, a fraction d_h of seniors leaves the market. Eventually, the number of seniors falls below the level required to meet the target MN. Once this occurs, seniors become scarce again. Their wages jump to the new steady-state level, and firms resume hiring juniors to complete the remaining tasks.

Proof of Proposition C.1. Whenever there is an oversupply of seniors ($N_t^h > MN/\gamma_{AI}^h$), the competition among seniors drives the wage of seniors down to b. If the firm still employs juniors, the wage should be strictly less than b (otherwise, the no-arbitrage condition is violated), in which case juniors will not join. As a result, firms do not hire juniors. With no new juniors, the senior stock is strictly decreasing over time $N_{t+1}^h = (1 - d_h)N_t^h$ when the oversupply persists.

We first determine the maximum duration workers are willing to wait for the oversupply of seniors to disappear (where the number of seniors is smaller than MN/γ_{AI}^h). Let $v_1^h(\tau)$ denote the value of a senior at $t = 1$, conditional on the market clearing at period $\tau > 1$. For all periods $t < \tau$, the wage is b. From period τ onward, the wage rises to the new steady state $w_h^*(\gamma_{AI}^h)$. The value function can be written recursively:

$$v_1^h(\tau) = \sum_{k=0}^{\tau-2} [\delta(1 - d_h)]^k (b + \delta d_h \underline{v}) + [\delta(1 - d_h)]^{\tau-1} v_{new}^{h*}.$$

Since $b < (1 - \delta)\underline{v}$, the flow utility during waiting is lower than the outside option. Consequently, $v_1^h(\tau)$ is strictly decreasing in the waiting time τ. Workers will only remain in the market if the value of doing so exceeds their outside option $\underline{v}$. We define T_{max} as the maximum tolerable duration of waiting:

$$T_{max} \equiv \max\{\tau \in \mathbb{N} \mid v_1^h(\tau) \geq \underline{v}\}.$$

Any expected wait longer than T_{max} drives the value of a senior job below $\underline{v}$, causing an immediate exit. This duration limit imposes a limit on the initial stock of seniors. During the waiting, the stock decays only through attrition: $N_{t+1}^h = (1 - d_h)N_t^h$. For the market to clear within T_{max} periods, the initial stock must decay to the target level MN/γ_{AI}^h by period

T_{max}. This defines the capacity threshold N_{max}:

$$N_{max} \equiv \frac{MN}{\gamma_{AI}^h}(1-d_h)^{-(T_{max}-1)}.$$

If the initial stock N_0^{h*} exceeds N_{max}, senior workers anticipate this and exit immediately until the stock falls to N_{max} at $t=1$. And we redefine the stock of the seniors at $t=1$ to be $N_1^h = N_{max}$ and define the waiting time $T = T_{max}$.

If $N_0^{h*} \leq N_{max}$, then all of them will stay, so $N_1^h = N_0^{h*}$. The waiting time T is the minimum τ such that $N_1^h(1-d_h)^{\tau-1} < MN/\gamma_{AI}^h$.

At period T, and given the stability constraint $S' > -1$, the transition follows Proposition 2. ■

C.2 A Gibbons and Waldman (1999) Framework

Production requires the completion of N tasks. We now model the firm as consisting of a job ladder with two levels, denoted by $j \in \{1,2\}$. Job 1 is equipped with basic production technology, while job 2 is equipped with advanced production technology that requires higher human capital.

Workers differ in their effective ability. Let $\eta_{it} \in \{0,1\}$ denote the effective ability of worker i in period t. A worker is either a junior with ability $\eta = 0$ or a senior with ability $\eta = 1$. In each period, a junior acquires the skills to become a senior with probability $q \in (0,1]$.

Following Gibbons and Waldman (1999*b*), we model the output of worker i on job j as a linear function of this ability:

$$y_{ijt} = d_j + c_j\eta_{it}.$$

For job 1, we set $d_1 = \gamma^l$ and $c_1 = 0$; for job 2, we set $d_2 = 0$ and $c_2 = \gamma^h$.

These parameters imply that job 1 yields γ^l regardless of the worker's ability, while job 2 yields γ^h for a senior but zero for a junior. Because of this difference, the efficient assignment rule places juniors in job 1 and seniors in job 2. This assignment rule yields the firm structure in our main model.