Received: 12 January 2021 | Revised: 4 May 2021 | Accepted: 24 May 2021

DOI: 10.1111/jems.12439

ORIGINAL ARTICLE

Journal of Economics & Management Strategy WILEY

Too big to succeed? Overstaffing in firms

Hans K. Hvide[1,2] | **Yanren Zhang**[3]

[1]Department of Economics, University of Bergen, Bergen, Norway

[2]Department of Economics, University of Aberdeen, Aberdeen, UK

[3]Fudan Development Institute, Fudan University, Shanghai, China

Correspondence
Yanren Zhang, Fudan Development Institute, Fudan University, Shanghai, China.
Email: zhangyanren@gmail.com

Abstract

Overstaffing appears to be a source of significant inefficiencies in organizations, but there is little economic theory that informs us why. We extend the canonical Lazear–Rosen tournament model to a dynamic setting that yields overstaffing at the managerial level. Overstaffing can be optimal in first best, without moral hazard, if the redundant manager gains experience and increases the firm's future productivity. In second best, overstaffing can be a way to provide incentives to young workers without "overpaying" middle-aged workers, a point that is illustrated with several examples from real world organizations. The model may offer some independent interest by integrating a generational structure into a tournament model.

> *An[other] important problem with promotion-based reward systems is that they require organizational growth to feed the reward system* (Baker et al., 1988, p. 600).

1 | INTRODUCTION

A frequent complaint about bureaucracies, both in the private and in the public sector, is that they waste resources by employing too many people. This is not a new complaint; Gibbon (1797) states that under the Roman emperor Diocletian, "The number of ministers, of magistrates, of officers, and of servants, who filled the different departments of the state, was multiplied [.]" Secretary of State Madeleine Albright stated in 1996 that "The U.N. bureaucracy has grown to elephantine proportions. Now that the Cold War is over, we are asking that elephant to do gymnastics." More recently, the seemingly excessive size of the European Union bureaucracy was an effective argument for the Brexit campaign in the United Kingdom.

By overstaffing we mean employing individuals that do not contribute to the organization's productivity in a static sense. Although overstaffing seems to be a pervasive empirical phenomenon (we provide more examples in Supporting Information Appendix A), there is little economic theory to tell us why it occurs. For example, neither the Lazear and Oyer (2012) survey of Personnel Economics nor the Waldman (2012) survey of internal labor markets mentions overstaffing. One way to explain overstaffing comes from Williamson (1964), who argues that managers derive utility from increasing the number of subordinates. Niskanen (1968), applies this idea to managers in the public sector, and Jensen (1986), to executive officers in publicly listed firms. The Williamson tradition explains overstaffing by managers extracting rents through empire-building, but does not explain where the rents come from or why they are extracted in this manner, rather than through, for example, a higher salary.

Can overstaffing occur in organizations, public or private, that attempt to motivate its employees in a cost-efficient manner? If so, why does the organization not avoid overstaffing by firing (or avoid hiring) surplus individuals? We build on the intuition of Baker et al. (1988) in the opening quote to analyze the trade-off between incentives and

allocation in a dynamic extension of the canonical Lazear and Rosen (1981) tournament model. Overstaffing in second best occurs as an optimal way to incentivize young workers to provide effort: positions higher up in the hierarchy that appear redundant serve the purpose of motivating lower-level employees.

The motivation for using the tournament model comes from empirical evidence suggesting that wages predominantly grow through job changes rather than within job grades. Consequently, incentives within organizations seem rely heavily from the prospect of future promotions (e.g., Baker et al., 1994a, 1994b; Lazear, 1991; Medoff & Abraham, 1980). Furthermore, evidence suggests that promotions are determined by workers' relative rather than absolute performance (De Varo, 2006). Workers therefore often compete with other workers for a limited number of higher-level jobs.

In the model, firms are long-lived and have two layers: Worker and managerial. Individuals live for three periods; they are young, middle-aged or old. In each period, two workers engage in a tournament, and the winner gets promoted to manager, with an associated pay increase. Workers may be promoted to the managerial level either after one period, after two periods, or not at all. Promoted workers may be managers for one period or for two periods. The latter is only possible if promoted when young; middle-aged individuals turn old and retire after one period at the managerial level. Only one manager is needed for production. Overstaffing occurs in the case that the incumbent manager has time left of his tenure when a worker is promoted, so that two individuals are employed at the managerial level even if only one of them contributes to contemporaneous productivity. The hiring and firing policy of the firm is endogenous; a simple way of avoiding overstaffing is to let individuals be managers for one period only.

The model isolates two reasons why a cost-efficient organization may choose to overstaff. The first is human capital acquisition at the managerial level, that is, managers are more productive in the second period of managerial tenure. If such acquisition is strong then even in first best, when effort provision is contractible, the firm would have two individuals employed at the managerial level. Under moral hazard, that is, when effort provision is not contractible, the firm introduces a stick and a carrot to incentivize workers. The stick is to reduce job security by using an up-or-out policy. The carrot is to lengthen tenure at the managerial level to two periods, which creates overstaffing. In other words, overstaffing incentivizes young workers to exert effort. The reason why lengthening tenure at the managerial level (i.e., overstaffing) is chosen as carrot, rather than paying a higher wage for a shorter period to managers and then firing them, is demography: overstaffing avoids the cost of overpaying middle-aged workers that are promoted. Thus overstaffing in our model emerges as a way to deal with young and middle-aged workers differently.

In symmetric one-shot tournaments, firms cannot benefit from falsely reporting performance rankings and discriminating against certain types of workers, because the wage bill is fixed (e.g., Bhattacharya & Guasch, 1988; Malcomson, 1984).[1] This property of one-shot tournaments does not carry over to our dynamic setting because the firm can have a preference for promoting a young worker (to increase future human capital at the managerial level) or for promoting a middle-aged worker (to avoid future overstaffing), regardless of relative performance. If the firm were to discriminate in this manner, the tournament breaks down as an incentive device. We analyze how the firm adjusts its personnel policy to make promotion based on relative performance credible. Somewhat surprisingly, the firm obtains credibility by reducing the amount of overstaffing (but wages increase, so profits drop).

The "efficient overstaffing" that occurs in second-best can shed light on puzzling features of real-life organizational charts. One case is Tesco Plc.[2] Tesco is a large UK-based retail company with more than 400.000 employees in almost 7000 stores. Individual employees' pay at Tesco is largely determined by rank so that pay increases—and thus incentives—stem from the prospect of being promoted to a higher rank. Tesco's organizational chart included positions called "people manager" and "compliance manager," positions whose job description indicate that they are junior to store managers but fill much the same role, such as recruitment, coaching, and other management functions.[3] These extra managerial positions had the benefit of providing incentives to lower-level employees (Grimshaw et al., 2002), and for years Tesco appeared to live well with these "extra" managers. In a recent move (apparently related to reduced demand due to increased UK competition), Tesco decided to cut the people manager and compliance manager positions from the organizational chart. The purpose of the delayering was to "improve efficiency and give line managers clearer accountability." Grimshaw et al. (2002) contains a broader perspective, by studying four large firms from the banking, urban city council, food retail, and telecommunication industries, that shredded a number of mid-management positions, similar to in Tesco. Through interviews with more than 100 employees, Grimshaw et al. (2002) document how this process had significant costs in weakening incentives and hurting morale.[4] Using survey evidence from firms located in three different countries, Littler et al. (2003) discuss the extent to which delayering is followed by "relayering." They find that relayering occurs in about 10% of their sample. These examples show that the trade-off between production efficiency and providing worker incentives, captured by our model, appears to be a quite general

conundrum for firms, and the solution not uncommonly being that the firm deliberately overstaffs at the managerial level to create incentives for lower-level employees.

The paper proceeds as follows. Below, we relate our work to the prior theoretical literature. Section 2 outlines the basic economic environment and derives the states and surplus generated by the firm. Section 3 analyzes the optimal personnel (hiring/firing/wages) policies in first best. Section 4 analyzes how moral hazard from worker's side can lead to second best overstaffing. Section 5 considers moral hazard from the firm's side, that is, inability to commit to a promotion rule, and Section 6 concludes.

1.1 | Related literature

The paper connects to several ongoing debates. In work based on Williamson (1964), organizations get bloated due to managers having a preferences for expanding the size of their empire; profits would go up if some employees are shredded. A notion of inefficient overstaffing is also an element of Acemoglu et al. (2011), where elites may prefer to hire excessive bureaucrats to get more votes, and in models of labor unions (e.g., McDonald & Solow, 1981, 1985; Oswald, 1985), where overstaffing occurs due to unions having a preference over employment level. Relative to the extant literature, our contribution is to use tools from personnel economics to show that overstaffing can occur with a cost-efficient organization that employs individuals with standard preferences. Moreover, we identify conditions under which overstaffing is more and less likely to occur.

The central trade-off of our paper—between productive efficiency (avoiding having surplus managers) and creation of incentives (having surplus managers)—has been studied by Ke et al. (2018) in a Shapiro and Stiglitz (1984) efficiency wage type of framework with homogenous workers. They show (Corollary 2 in their paper) that firms opt to have a higher fraction of managers to workers than what is technically efficient. Although not an overstaffing result (which would be that the sum of workers and managers would be too large), the intuition of their result is similar to ours: to motivate lower-level employees, the firm sacrifices static productivity.[5]

The seminal Lazear and Rosen (1981) has spawned a large literature that analyzes promotion-based incentives. Examples of the topics that have been explored by using tournament models are handicapping (e.g., Imhoff & Krä-kel, 2016), R&D competitions (e.g., Boudreau et al., 2016), Chief Executive Officer remuneration (e.g., Rosen, 1986), sabotage (e.g., Lazear, 1999), risk taking (e.g., Hvide, 2002), multistage tournaments (e.g., Goltsman & Mukherjee, 2011), and relational contracts (e.g., Zabojnik, 2012). Somewhat surprisingly, the tournament literature has not analyzed overstaffing before. Our contribution is to extend tournament theory to a dynamic setting where worker demography counts, and to show that such an extension can produce overstaffing as an equilibrium outcome.

Finally, a large literature considers taste-based discrimination in the workplace (Becker, 1957; Darity & Mason, 1998; Neumark & Stock, 1999; Ross, 2003; Yinger, 1998). For example, according to the glass ceiling hypothesis, females are less often promoted than males because of a preference for male promotion among the higher levels of the firm's hierarchy. Schotter and Weigelt (1992) analyze tournaments with discrimination and show that discrimination leads to lower effort among all (otherwise homogenous) agents. Schotter and Weigelt (1992) consider a setting with assumed taste-based discrimination, by contrast we develop a tournament model where discrimination (a preference for promoting young or middle-aged workers) arises endogenously. We moreover show that this preference can be partially overcome, and incentives restored.

2 | MODEL

2.1 | Setting

2.1.1 | Organization

There is one infinitely lived organization with discount factor β. Time is discrete and the net present value at time t equals

$$\Pi_t = \sum_{\tau=t}^{\infty} \beta^{\tau-t} \pi_\tau, \tag{1}$$

where π_τ is period τ surplus (or profit). For a private-sector firm this is simply the net operating income and for a public-sector organization it is the difference between the (consumer) surplus generated by the citizens for the services offered by the organization minus the operating costs. In the following we use "firm" and "organization" interchangeably and intend it to cover both private-sector firms and public organizations. The organization maximizes its steady-state net present value by choosing a wage policy and a hiring/firing policy, to be described in more detail below. In each period, production requires two individuals employed at the worker level and one individual employed at the managerial level (minimum staffing). By "overstaffing," we mean that two individuals are employed at the managerial level, an issue we return to below. Suppressing time subscripts, the output of worker i is denoted by $y_i \in \{0, 1\}$, where $i = 1, 2$. Output is additive so that the total output equals $y_1 + y_2$. In the following, we drop the worker i subscript except when necessary.

2.1.2 | Individuals

Each individual is risk-neutral and lives for three periods. We denote young individuals by type I, middle-aged individuals by type II, and old individuals by type III. An individual exerts effort $e \in \{0, 1\}$ with cost of effort equal to $e\psi$ (here $\psi > 0$). Effort increases output in the first-order stochastic dominance sense, that is, $\Pr(y = 1 \mid e = 1) = q^H > \Pr(y = 1 \mid e = 0) = q^L$. It follows that $\mathbf{E}(y) = 2q^H$ if both workers choose $e = 1$. We assume that ψ is sufficiently small relative to $q^H - q^L$ so that the firm always prefers $e = 1$ in equilibrium.

Let w_t denote the wage payment in period t. Each individual has a separable utility function over monetary transfer and effort:

$$U = \sum_{t=0}^{2-\tau} \beta^t \left(w_t - e_t \psi \right), \tag{2}$$

where τ is the individual's age, that is, a young individual has $\tau = 0$, a middle-aged individual has $\tau = 1$, and an old individual has $\tau = 2$. In addition, a worker who switches firms bears a displacement cost $\delta > 0$. An alternative interpretation of δ is as the loss of firm-specific capital if the workers moves. Individuals are risk neutral but have limited liability so we confine attention to wages being nonnegative.[6] For simplicity, we normalize the per-period salary at the worker level to zero, and assume that the reservation utility for each worker is zero. In other words, the moral hazard at the worker level cannot be eliminated unless the firm can promote some workers to highly paid positions.

2.1.3 | Tournament

The firm can observe the ranking of worker output (y_1, y_2) and runs a promotion tournament in each period to mitigate the moral hazard. We assume that individuals with the same amount of experience in a given position are paid the same. Specifically, the per-period salary at the managerial level depends on job tenure, that is, a manager will obtain V_1 in his first period of tenure and V_2 in his second period of tenure (if applicable).

Let us denote the present value from being promoted by $W(V_1, V_2)$, where W depends on V_1 and V_2. Conditional on W, each worker chooses effort level e_i, individual output y_i is realized, and the firm ranks the workers. In the event of a tie, that is, $y_1 = y_2$, the toss of a fair coin determines the winner. Thus, if $e_1 = e_2$, then the probability of winning is $1/2$ for both workers; if $e_1 = 1, e_2 = 0$, then worker 1 outperforms worker 2 with probability $1/2 + (q^H - q^L)/2$, where $(q^H - q^L)/2 = \Delta p \in (0, 1/2)$ is the increase in the probability of winning from exerting effort if the other worker does not exert effort. Note that Δp measures the informativeness of the signal y about e. The larger the Δp, the higher the quality of monitoring system.

2.1.4 | Personnel policies

After the outcome of tournament is announced, the winner will be promoted while the loser will be fired or rehired. Hence, in each period, the firm has one or two slots available at the worker level and fills these by hiring young workers in the labor market. The firm's *personnel policy* specifies (i) whether a worker that loses the promotion tournament gets fired or rehired

(up-or-out vs. up-or-stay), (ii) whether to rehire as manager an individual who has already been manager for one period (managerial tenure length), and (iii) managerial wages V_1 and V_2. Sometimes we refer to (i)–(ii) as the firm's hiring/firing policy and (iii) as the firm's wage policy. We assume the personnel policies do not depend on the state of the firm, that is, (i)–(iii) are set permanently at time zero. Table 1 summarizes the possible personnel policies.

To make a simple policy space, we ignore two possibilities: (i) the firm may recruit one or two middle-aged workers in the labor market (old workers must retire in the next period and are impossible to incentivize, and will thus never be hired); (ii) the firm may apply the personnel policies that are state-dependent. In the paper's working paper version,[7] we consider the case in which the firm may recruit middle-aged workers and has access to state-dependent policies. We demonstrate that the main conclusions continue to hold in a larger policy space, because (i) the firm prefers to recruit young workers rather than middle-aged ones in the labor market and (ii) any state-dependent policy is weakly dominated by a state-independent one.

2.1.5 | Managerial human capital

We assume that human capital increases with job seniority at the managerial level. Specifically, if rehired by the original firm, a manager increases production by $\theta \geq 0$, so that with an experienced manager in place, the firm's total production equals $y_1 + y_2 + \theta$. By contrast, if the manager leaves for an outside firm, he increases production by $\kappa \geq 0$. The productivity improvement due to managerial experience is crucial to understand the overstaffing in first best but has nothing to do with the distortions in second best.

2.1.6 | Overstaffing

By the firm being overstaffed in period t, we mean that it engages more persons at the managerial level than what would maximize short-term profits. Overstaffing does not imply that the firm is wasting resources. We show in Proposition 1 that, in a dynamic setting, it will be optimal for the firm (in a first best sense) to be overstaffed if θ is sufficiently large. The reason is that if θ is large it pays off to have an unproductive middle-aged manager who will become productive when old. The focus of our analysis will be on situations where the firm is overstaffed even when θ is low. As shown in Proposition 2, such overstaffing (in a second best sense) will occur as an adaptation to employees' moral hazard.

To make the model interesting, we need a cost of overstaffing managers. Assume that there exists an outside manager market, where numerous firms compete for general management skills such that their market value precisely equals their productivity κ. In this situation, if not being rehired by the original firm, an experienced manager can still earn κ in the outside market.[8] Symmetrically, if hiring an experienced manager in the market, the firm will obtain the value of general human capital κ, and pay him the market salary κ. Since the value of hiring an external manager is zero (complete rent dissipation), rehiring internal managers becomes the only profitable way to acquire manager-level human capitals.

2.1.7 | Equilibrium

The dynamic game has two stages. In the first stage, the firm selects an element from the set {F-F,F-R,R-F,R-R} in Table 1 and an associated wage profile V_1, V_2. In the second stage, given the firm's choice of personnel policy, overlapping generations of finitely-lived workers play an infinitely repeated tournament. We focus on noncooperative

TABLE 1 Hiring/firing policies and staffing levels

	Policy F-F	Policy F-R	Policy R-F	Policy R-R
Nonpromoted worker I	Fired	Fired	Rehired	Rehired
Manager II	Fired	Rehired	Fired	Rehired
Staffing level	Minimum	Over-	Minimum	Cyclical-
	Staffing	Staffing	Staffing	Overstaffing

Note: The last row illustrates the corresponding staffing levels in steady state, which will be discussed in the next subsection.

behaviors (rule out collusion among workers) by restricting the workers to applying Markov strategies, that is, the workers' strategies depend only on the current state rather than game history.[9] A profile of strategies is a perfect equilibrium if and only if (*i*) in the first stage, the firm selects a personnel policy that maximize its net present value, and (*ii*) in the second stage, each worker i decides whether to participate in the tournament, and, if he participates, selects an effort level to maximize his utility given the effort level of the other worker. Moreover, individuals have the option to leave the firm after one or two periods.

2.2 | Firm profits

We now describe the profits of the firm under the hiring/firing policies described in Table 1. We summarize the type of overstaffing that occurs under each policy in the bottom row of Table 1. Recall that the firm launches a new tournament each period. We define period t profit as total production minus the present value of future promised wage payments. For example, for the period t tournament under Policy F-F, the future wage payment is V_1. As this cost is realized in period $t+1$, we need to discount it by β, so that the period t profit is $2q^H - \beta V_1$ rather than $2q^H - V_1$.

2.2.1 | Policy F-F

Under Policy F-F, the firm fires the nonpromoted worker (up-or-out rule) and does not rehire any manager II. Suppose the firm has two worker Is and one manager II in period $t = 0$. In period $t = 1$, the initial manager is fired and then replaced by the winning worker. Moreover, the losing worker is fired, and two new worker Is are recruited to fill the vacancies. therefore, there are two young workers and one middle-aged manager in steady state, which implies minimum staffing of the firm and the period profit $2q^H - \beta V_1$. Again the first term is the worker's expected output in period t and the second term is the present value of the promised wage payment in period $t+1$. Managerial human capital acquisition θ is not relevant because a promoted worker will retire as soon as his one-period tenure expires.

2.2.2 | Policy F-R

Under Policy F-R, the firm promotes one worker I and fires the nonpromoted worker (up-or-out rule). Moreover it rehires a manager II. So, in all periods $t \geq 2$, the firm is overstaffed because the firm has two individuals (one middle-aged and one old) at the managerial level, in addition to two worker Is. The profit generated in period t (total production minus the present value of promised wage payments) becomes

$$\pi = 2q^H + \beta^2\theta - \beta V_1 - \beta^2 V_2. \tag{3}$$

The first term is the workers' expected output in period t, the second term is the present value of the output from experienced manager III (who wins in period t) in period $t+2$, the third term is the present value of the promised wage payment in period $t+1$, and the fourth term is the present value of the promised wage payment in period $t+2$.

2.2.3 | Policy R-F

Policy R-F is the opposite of Policy F-R in that it rehires the nonpromoted worker (up-or-stay rule) but fires the manager II. There will therefore be no overstaffing under Policy R-F. The firm's organizational structure under Policy R-F is more complicated than those under Policies F-F and F-R because when losing worker Is are rehired, the promotion tournament becomes asymmetric, in that workers of different age compete. In other words, under Policy R-F, the state of the firm is determined by a stochastic process where the firm transitions between asymmetric and symmetric states with a positive probability. Nevertheless, the period profit under Policy R-F is irrelevant to the state because the wage bill is fixed and equal to V_1 in each period, which implies the firm has period profit

$$\pi = 2q^H - \beta V_1. \tag{4}$$

2.2.4 | Policy R-R

Under Policy R-R, the firm retains the nonpromoted worker (up-or-stay rule) and the manager II (if any). This means that overstaffing occurs whenever the incumbent manager is middle-aged. Similar to that under Policy R-F, the state of the firm is determined by a stochastic process where the firm transitions between asymmetric and symmetric states—asymmetric state meaning that two workers of different age compete—with a positive probability. Under Policy R-R, firm profits follow a stochastic process because the per-period profits depend on the state the firm. The net present value of the firm under Policy R-R is derived in Supporting Information Appendix B.

To illustrate the transition between states under Policy R-R, suppose that, in the initial period $t = 0$, the firm has two worker Is and one manager II. In period $t = 1$, the firm has two workers, one type I (newly hired) and one type II (the rehired loser of the $t = 0$ tournament), and two managers, one type II (newly promoted) and one type III (the initial manager). In consecutive periods, the firm has two possible configurations at the worker level: The promotion tournament played is either *symmetric* (a worker I against a worker I) or *asymmetric* (a worker I against a worker II). At the managerial level, the firm has four possible configurations: there is either one manager II, one manager III, one manager II and one manager III, or two manager IIIs.[10] As a result, there are eight substates of the firm in total, as illustrated in Table 2.

Note that overstaffing at the managerial level occurs in Substates X2, X4, Y2 and Y3. To explain how overstaffing can occur, consider the case where a worker I wins in period t. He will then be a manager II with a two-period tenure (from period $t + 1$ to $t + 2$). In period $t + 1$, the firm starts a new round of tournament, and the winner becomes a new manager in period $t + 2$. In other words, there will be two individuals at the managerial level in period $t + 2$ if a worker I wins in period t. In contrast, if a worker II wins in period t, he will have a one-period tenure and then retire before a new manager moves into the office in period $t + 2$. Hence, overstaffing will not occur.[11]

3 | FIRST BEST

We start out with an analysis of first best, that is, when worker effort is contractible. In first best, the firm can commit to firing any worker selecting $e = 0$. Assume each individual will prefer $e = 1$ if he is indifferent. Then, on the equilibrium path, each worker will select $e = 1$ and obtain his reservation utility zero. In other words, the firm in first best can restrict its attention to participation (IR) constraints and ignore incentive-compatible (IC) constraints.

In addition to the worker ex ante participation (IR) constraint (zero reservation utility), the firm faces a set of interim participation constraints for managers, which ensure that no manager wishes to leave the firm. It is easy to know that a promoted worker is willing to accept the managerial job in his first period of managerial tenure regardless of hiring/firing policies. This is because outside firms pay zero for a winner in an internal promotion tournament, but κ for an experienced manager. To acquire managerial experiences, the promoted worker wishes to stay in his first period of tenure. In his second period of tenure (if rehired), he accepts the managerial job as an experienced manager if and only if $V_2 \geq \kappa$.

To make the comparison with second best interesting, we assume that the firm randomly promotes one worker to managerial level each period (this is equivalent to not allowing discrimination on age in the promotion decision).

TABLE 2 Possible substates of firm under policy R-R

Substate of firm	Workers	Manager(s)
X1	I + I	II
X2	I + I	II + III
X3	I + I	III
X4	I + I	III + III
Y1	I + II	III
Y2	I + II	II + III
Y3	I + II	III + III
Y4	I + II	II

Therefore, the firm has the same policy options as described in Table 1.[12] Figure 1 represents the optimal policies in the (θ, δ) space.

The solid blue lines split the (θ, δ) space into subsets. For $\theta \geq \kappa$ (to the right of the vertical line), the firm overstaffs and retains managers for two periods (Policies F-R or R-R) and for $\theta \leq \kappa$, the firm retains managers for only one period (Policies F-F or R-F) and is not overstaffed. For small worker displacement cost δ (below the horizontal line, that is, F-F/R-F and F-R/R-R partition lines) the firm keeps workers for only one period, that is, an up-or-out policy (Policies F-F or F-R), and for large δ the firm applies an up-or-stay policy (Policies R-F or R-R). We summarize these results in the following proposition.

Proposition 1. *In first best, the firm*

(i) rehires managers if and only if $\theta \geq \kappa$,
(ii) rehires nonpromoted workers if and only if δ is sufficiently large.

Proof. See Supporting Information Appendix C. □

Why does a given policy dominate others in a given parameter range? To answer this, we need to know how θ, δ and κ affect the optimal hiring/firing policy. First, rehiring experienced managers (Policies F-R or R-R) becomes more profitable when the value of management experience θ is sufficiently large. Second, recall that workers bear the displacement cost δ. If the firm offers job security, on the one hand, workers might be heterogenous, which increases the total employment cost; on the other hand, the payment to induce workers to participate would decrease. When δ is sufficiently large, the latter effect dominates the former one, and hence the firm prefers to rehire nonpromoted workers (Policies R-F or R-R). Accordingly, the firm should apply Policy F-F if both θ and δ are small, Policy R-R if both are large, Policy R-F if θ is small but δ is large, and Policy F-R if θ is large but δ is small, as shown in Figure 1. Finally, recall κ represents the outside option of experienced managers. So, with the increase of κ, the cost of rehiring managers would rise, which strengthens the relative advantage of Policies F-F or R-F over F-R or R-R.

Overstaffing in first best (which occurs for $\theta \geq \kappa$ only) is of two different types. For low δ (i.e., under Policy F-R), the firm applies an up-or-out policy, fires nonpromoted workers, and is overstaffed in all periods. Here the firm runs a "management training program" by in every period inducting a new worker to manager level, and letting that new manager stay on for two periods, in the first managerial period (when middle-aged) jointly with an older manager and in the second period (when old) jointly with a middle-aged manager recently promoted. When worker displacement cost δ is high, on the other hand, the firm offers job security to workers, that is, does not fire nonpromoted workers, and promotes a middle-aged worker to managerial level on average every second period. Overstaffing will occur two periods after a young worker is promoted, that is, when that person is an old manager. For high δ (i.e., under Policy R-R), overstaffing occurs on average two thirds of the periods. Because the probability that the firm has an experienced manager is 2/3 under Policy R-R and 1 under Policy F-R, Policy F-R becomes more profitable relative to Policy R-R, as θ increases. This explains why the partition line between Policies F-R and R-R is upward sloping.

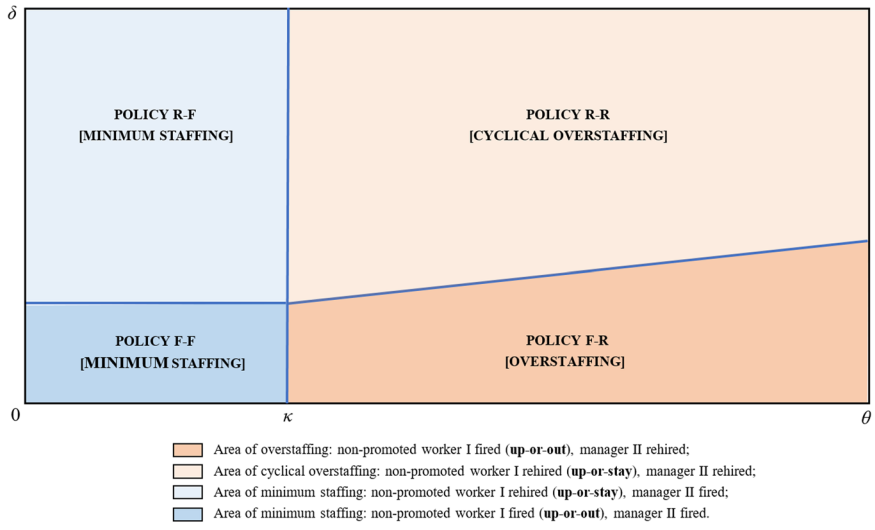

FIGURE 1 First-best hiring/firing policy [Color figure can be viewed at wileyonlinelibrary.com]

4 | SECOND BEST

Under first best, overstaffing occurs because of human capital acquisition at the managerial level. We now turn to the analysis under moral hazard, and show that second best overstaffing can occur, meaning overstaffing for values of (θ, δ) where overstaffing does not occur in first best.

When effort is noncontractible, the firm sets (V_1, V_2) so that each worker prefers to exert effort. Workers will have incentives to exert effort if,

$$\Delta p W_i(V_1, V_2) \geq \psi, \tag{5}$$

or

$$W_i\left(V_1, V_2\right) \geq \frac{\psi}{\Delta p}. \tag{6}$$

Here, Δp is the increase in win probability from exerting effort, $W_i(V_1, V_2)$ is the associated increase in utility (which will depend on the firm's hiring/firing policy) and ψ is the cost of effort. This is a standard type of incentive-compatibility (IC) condition from the tournament literature.[13] The firm then solves the following problem:

$$\max_{\{P, V_1, V_2\}} \Pi \quad \text{s.t. IR constraints and IC constraints.} \tag{7}$$

where $P \in \left\{F\text{-}F, F\text{-}R, R\text{-}F, R\text{-}R\right\}$. In other words the firm picks the personnel policy that maximizes steady-state net present value subject to the incentive compatibility and participation constraints. Figure 2 illustrates the distortion of hiring/firing policies when moving from first to second best, when κ is low.

Compared to the first best, the region where up-or-out is used (nonpromoted workers are fired) expands, that is, the horizontal line shifts up. Job security for workers weakens the incentives of the tournament due to the "second chance" effect—the worker may be promoted later and therefore has weaker incentives to exert effort now—and the firm uses up-or-out as a stick to incentivize workers in this area.

Beside using a stick to incentivize workers, second best also involves a carrot: compared to the first best, the area with longer tenure for managers (more overstaffing) expands, that is, the vertical line shifts to the left (for sufficiently large δ). This increases the value of being promoted for young workers. The logic is simple. Suppose managers are never rehired. Worker Is have stronger incentives to shirk than worker IIs, because they have a second chance to be promoted if they fail in the first round. To incentivize worker Is without paying too much to worker IIs, the firm defers the payment to worker Is by allowing them to be managers for two periods upon winning the tournament.

An example can illustrate second-best overstaffing. Suppose that worker Is are not willing to invest high effort unless paid at least \$1000 if promoted, while worker IIs are willing to invest high effort for \$500. Recall that managers

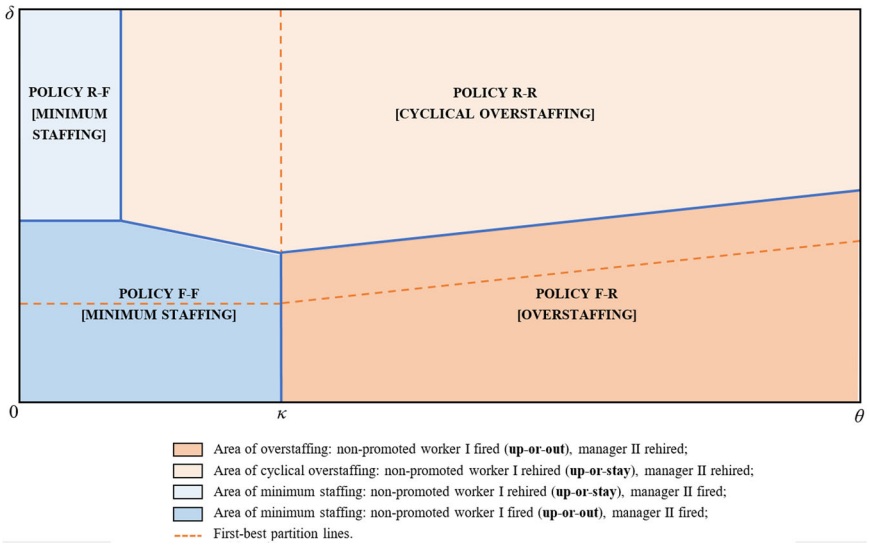

FIGURE 2 Second-best hiring/firing policy (low κ) [Color figure can be viewed at wileyonlinelibrary.com]

with same job seniority are equally paid. To induce high efforts from both types of workers, the firm should offer $V_1 = \$1000$ to promoted workers in their one-period manager tenure. However, the firm can make its compensation scheme more flexible by extending the tenure of manager IIs. In particular, neglecting discounting, the firm can set $V_1 = V_2 = \$500$, which preserves the incentive compatibility (a winning worker I obtains $V_1 + V_2 = \$1000$, and a winning worker II obtains $V_1 = \$500$) at lower cost (because the expected incentive cost becomes $\$1000 \times 1/2 + \$500 \times 1/2 = \$750$). In Figure 2, the partition line between Policies R-F and R-R is shifted left, which leads to a bulge in the area covered by Policy R-R.

The bulge implies that it is possible for the firm to replace Policy R-F with R-R even when the value of managerial experience θ is low, which raises two issues. First, when $\theta < \kappa$, unlike in first best where the choice is only between Policies F-F and R-F, the firm might choose between Policies F-F and R-R. Graphically, a new partition line between Policies F-F and R-R appears in Figure 2. Recall that the slope of the partition line between two policies indicates how the relative profitability of the two policies changes with θ. So, the new downward sloping partition line shows that the firm is more inclined to replace Policy F-F with R-R as θ increases. The logic is simple. The probability that the firm has an experienced manager is 2/3 under Policy R-R but is zero under Policy F-F. As a result, Policy R-R becomes more profitable relative to Policy F-F with the increase of θ.

Second, we can note that in the Policy R-R area, left of κ, a manager in his second-period tenure is paid above both the market wage ($V_2 > \kappa$) and his marginal product ($V_2 > \theta$). Intuitively, worker Is are willing to exert high effort only if managers earn high V_1 or V_2 in the firm, or alternatively earn high κ in the market. Given κ is low, the firm has to increase V_1 or V_2 to preserve incentive compatibility. The reason for raising V_2 above κ, rather than increasing V_1, should be familiar by now: Increasing V_2 creates incentives for worker Is and avoids overpaying worker IIs. Since the firm rehiring managers would not only reap the product θ but also save the total cost of incentive, raising V_2 above θ might be profitable.

When κ is intermediate, worker Is' incentive compatibility constraint is slack because being promoted pays at least κ when old. In that case, there is no need to "overpay" rehired managers (i.e., set $V_2 > \kappa$) to incentivize worker Is, so the vertical line is the same in second best as in first best. The firm still adopts an up-or-out policy too often compared to first best, again to incentivize workers. In other words, the horizontal line in Figure 2 shifts up relative to first best.

Finally, in the relatively uninteresting case where κ is high, the outside wage for an experienced manager will be sufficient to incentivize workers and the firm will be indifferent between rehiring losing workers or not. In the case where κ is so high that the firm running a "management training program" can set V_1=0, Policy F-F is equivalent to Policy R-F because the firm's wage bill equals zero for both policies. The following proposition summarizes our results.

Proposition 2. *Compared to first best, in second best there is, for sufficiently low κ,*

(i) less job security at the worker level (more up-or-out).
(ii) longer tenure at the managerial level,
(iii) more overstaffing.

Proof. See Supporting Information Appendix D. □

We have shown that second best overstaffing occurs in a model where pay is determined by relative performance only and where the firm's personnel policy is constant across all states. This type of overstaffing, which occurs in the area left of the vertical line in Figure 2, is a way to incentivize young workers to exert effort. The reason why such deferred payment is optimal, rather than paying a higher wage for a shorter period to managers (and then fire them), is demography; overstaffing is a way to incentivize young workers without overpaying middle-aged workers. A natural question is whether an age-dependent pay does away with such overstaffing, that is, what if initial wage as manager, V_1, can be made conditional on the age when promoted. Assume that the firm can pay $V_{1,I}$ if promoted as young and $V_{1,II}$ if promoted as middle-aged person. Then the firm would set $V_{1,I} > V_{1,II}$ and fire managers after one period (i.e., replacing Policy R-R with Policy R-F), thus avoiding overstaffing in the area left of the vertical line in Figure 2. Moreover, the firm would provide more job security, which may alleviate the problem of overstaffing in the area below of the horizontal line in Figure 2. See Supporting Information Appendix E for details.

The assumption that V_1 is independent of age seems reasonable for two reasons; first it aligns with fairness norms (equal pay for the same job) and the apparent wage structure of many firms. It is also consistent with laws that prohibit age-based discrimination, such as the Age Discrimination in Employment Act (ADEA) in the United States, which

prohibits age being a factor determining pay for workers above 40 years old.[14] A possible counter argument is that the firm could potentially reduce overstaffing by conditioning V_1 on how long the promoted worker has been employed in the firm, which is not illegal. However, this would involve paying more experienced individuals less due to the "second chance" effect for young workers, which appears unreasonable and hence is assumed away by the assumption that individuals with the same amount of experience in a given position are paid the same.

5 | TWO-SIDED MORAL HAZARD (THIRD BEST)

In static tournaments, firms cannot benefit from falsely revealing performance rankings because the effort of workers is sunk. In our setting, in contrast, whether a young or a middle-aged person is promoted affects future profits, and the firm may have an incentive to cheat ex post on the performance ranking. We now analyze the case where the firm can commit to a personnel policy, but cannot commit to promoting the top performer. We thus get a two-sided moral hazard problem. It is easy to find that, under Policies F-F, F-R, and R-F, the firm is indifferent as to which worker to promote and thus cannot benefit from cheating on the performance ranking. Specifically, under Policies F-F and F-R, the workers competing for promotion are identical, and the firm is indifferent about which worker to promote. The firm is also indifferent as to which worker to promote, young or middle-aged, under Policy R-F, because managers are not rehired. When the firm retains nonpromoted workers and rehires managers (Policy R-R), the lack of ability to commit bites.

Lemma 1. *Under Policy R-R, the firm has incentives to promote middle-aged workers if $\theta < V_2$, and has incentives to promote young workers if $\theta > V_2$. To ensure that merit-based promotion is self-enforcing, the firm sets*

$$V_2 = \theta. \tag{8}$$

Proof. See Supporting Information Appendix F. □

Lemma 1 shows that the firm can endogenously have a preference for discriminating against workers under Policy R-R. The intuition is straightforward. Suppose that the firm sets $V_2 > \theta$, that is, the wage for an experienced manager exceeds marginal productivity. Then, if a worker I is promoted the firm will bear a loss on this person when he becomes an experienced manager. To avoid this loss, the firm has an incentive to cheat on the performance ranking and instead promote a worker II. Likewise, if the firm sets $V_2 < \theta$, the firm will run a profit on experienced managers and have incentives to promote worker Is. To solve the commitment problem, the firm should equalize the payoffs from different promotion decisions by setting $V_2 = \theta$.

It is clear from the above remarks that, under Policies F-F, F-R, and R-F, firm profits are the same under second and third best, while the profits under Policy R-R decrease. Therefore the area where Policy R-R is chosen must shrink. In addition, if the firm chooses Policy R-R it must adjust V_2 according to Lemma 1. Formally, the firm solves the "third best" profit maximization problem,

$$\max_{\{P,V_1,V_2\}} \Pi \quad \text{s.t.} \quad \text{worker IR and IC constraints,} \quad \text{firm IC constraint.} \tag{9}$$

We now illustrate how the third best solution differs from the second best solution, that is, the impact of non-commitment, in the case where κ is low.

Figure 3 illustrates the distortion of hiring/firing policy due to moral hazard from firm's side. The bulge area on the interval $\theta < \kappa$ (from Figure 2) is eliminated because the firm can no longer credibly promise to pay $V_2 > \theta$, because $V_2 > \theta$ would imply that the firm would have a preference for promoting worker IIs, and thus violate the firm's IC constraint in Lemma 1. Since Policy R-R does not work anymore in this area, the firm adopts Policy R-F, which it also does in first best. In other words, the productive efficiency in this area is higher in third best than in second best (but the firm's profits decrease).

The second adaptation of the firm is to increase the tendency to use up-or-out, that is, the horizontal border shifts up relative to second best. The reason is that since the firm can no longer provide incentives to young workers by "overpaying" them as managers in second-period tenure, the firm instead strengthens incentives for young workers by adopting up-or-out to a greater extent. The alternative way of providing incentives, to increase V_1, would be more costly because this would induce payments also to middle-aged workers, whose incentive constraint is slack.

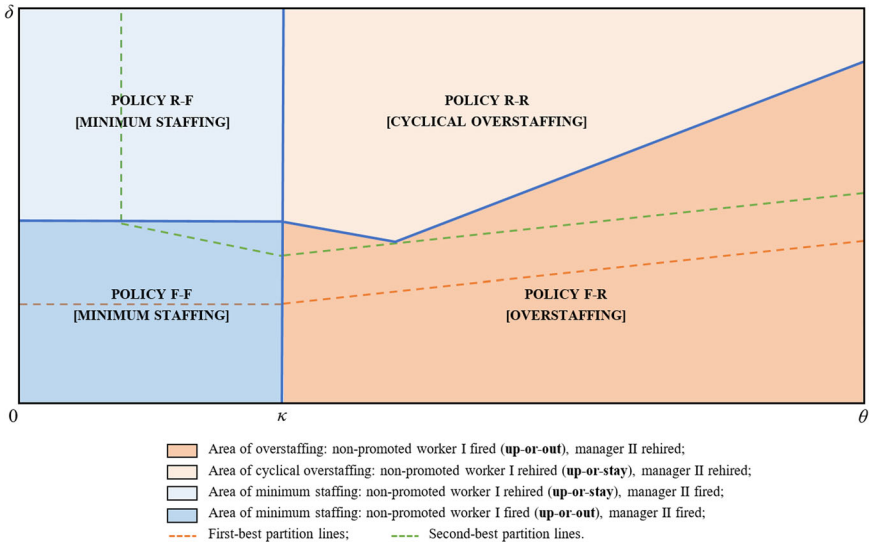

FIGURE 3 Third-best hiring/firing policy [Color figure can be viewed at wileyonlinelibrary.com]

Somewhat surprisingly, the regions of overstaffing in third best are exactly the same as in first best. This does not mean that allocations are efficient under third best: To the right of the vertical border in Figure 3, some of the area of cyclical overstaffing (Policy R-R) has been replaced by an area with constant overstaffing (Policy F-R), and therefore the frequency of overstaffing has increased (in first best there is overstaffing on average two out of three periods, while in third best there is overstaffing in every period). For the same reason, there is also more overstaffing in third best than in second best. For $\theta < \kappa$, the area of second best overstaffing has been eliminated in third best. There is therefore less this type of overstaffing in third best than in second best. The following proposition summarizes these results.

Proposition 3. *Compared to second best, if the promotion rule is noncontractible, for sufficiently low κ, there is*

(i) less job security at the worker level (more up-or-out),
(ii) a shorter tenure at the managerial level,
(iii) experienced managers are paid their marginal product ($V_2 = \theta$),
(iv) ambiguous effects on overstaffing.

Proof. See Supporting Information Appendix G. □

Of course, if the firm can establish a reputation for promoting the best performing workers, second best policies can be applied even in the noncommitment case. Generally speaking, in a repeated game with long-run and short-run players, a long-run player (firm) can be induced to live up to their promise by the prospect of future gains and punishment. Sustaining a reputation necessitates a large punishment if the firm deviates from a performance-based promotion rule and is caught. In the current context, it is not clear which party should enforce this punishment. For example, although it is possible to establish reputation-based equilibria where future workers refuse to be hired by the firm if the firm deviates from performance-based promotion, there will clearly be incentive issues for future workers to implement the punishment. Therefore, this model has little to say about whether the reputation mechanism is a more reliable way to ensure the credibility of a merit-based promotion rule than the wage adjustment we propose.

6 | CONCLUSION

Overstaffing appears to be an important source of inefficiencies in organizations, but has received little attention both from theorists and from empirical work. We show that an extension of the canonical Lazear and Rosen (1981) model can produce overstaffing as an equilibrium (efficient) outcome. The first reason is human capital acquisition at the managerial level, where the firm is running a "management training program" by having new managers work alongside more experienced managers. The second, more interesting, is that overstaffing serves to incentivize entry-level individuals to exert effort. The reason why such backloading of payment is optimal, rather than paying a higher

wage for a shorter period for managers (and then firing them), is demography; keeping managers for "too long," and thus overstaffing, incentivizes young workers to exert effort without overpaying middle-aged workers.

Our model highlights that overstaffing at the managerial level may be an efficient response by firms that use promotion-based incentives, and delineates conditions under which such overstaffing is more likely to occur. Although highly stylized, the model appears to capture quite well overstaffing in real-world firms. In the paper's working paper version, we consider three extensions of the basic model,[15] and show that our main findings are robust to concerns about (i) favoritism in asymmetric state, (ii) promotion rule based on fixed output standard, and (iii) a larger set of feasible personnel policies. Overall, we believe the model is useful in making the notion of overstaffing in promotion-based incentive systems, as highlighted by Baker et al. (1986), more precise. It may also form the basis of future theoretical and empirical work.

ENDNOTES

[1]Kahn and Huberman (1988) build a model where the firm has incentives to misreport worker performance in a nontournament setting. Carmichael (1983) and Tsoulouhas (1999) contain two-sided moral hazard models where the principal and agents work together as a team in a tournament setting. The only other paper we are aware of with an imperfect link between performance and promotions is by Fairburn and Malcomson (2001), who analyze a setting where workers can bribe middle managers to affect their performance rating. They show that this problem can be solved by linking middle managers' pay to division performance. In contrast, we analyze a setting where firms may deviate from merit-based promotion even in the absence of middle managers.

[2]The details from this example are taken from https://www.thegrocer.co.uk/hiring-and-firing/tesco-scrapping-over-1700-management-jobs/562462.article, and were also covered by the Guardian (https://www.theguardian.com/business/2018/jan/22/tesco-to-shed-1700-jobs-in-new-management-shake-up). General information about Tesco can be found at https://www.tescoplc.com/reports-and-policies/our-approach-to-reward/.

[3]A job description for people managers can be found at https://www.tesco-careers.com/jobdetails/101817.

[4]Grimshaw et al. (2002, p. 99) report that "Staff in each of the organizations expressed concern about these developments. [.] Managers [.] recognized the limited prospects for promotion opportunities and the implications this might have for morale. A store manager at Retailco explained to us that in the new delayered store: The promotional leap is too harsh for many general assistants... a lot of staff get frustrated with this leap... few staff feel they can do it." These sentiments appeared common in all the four firms studied by Grimshaw et al. (2002); at the banking firm, 40% of respondents in a staff survey considered "lack of opportunities to progress" as the worst feature of the firm. At some point, the paucity of promotion opportunities also became a concern for the higher management, a representative expressing that " the firm] now seeks to overcome retention problems by putting in place a clear career ladder...," or in other words a partial reversal of the reform.

[5]As a lesser point, turnover is assumed to be higher in the "managerial" task in the Ke et al. setting (this is what drives rents being higher in the "managerial" task and the possibility of using it to incentivize workers in the other task). This is in contrast to the interesting situations where some individuals stay on in the managerial job "forever," a situation we try to capture.

[6]With risk-neutral agents, moral hazard would be alleviated through the "selling-the-enterprise-to-the-agent" solution. In this paper, we assume limited liability/liquidity constraints so that no worker can buy the firm.

[7]See https://fddi.fudan.edu.cn/fddien/8c/0f/c19513a297999/page.htm.

[8]We abstract from displacement cost at the manager level because κ can be interpreted as the manager's gross salary minus displacement cost in the alternative occupation.

[9]A large literature explores how cooperative behavior can be sustained by a (non-Markov) subgame-perfect equilibrium, for example, Cremer (1986), and Kandori (1992).

[10]Note that the situation with two middle-aged individuals at the managerial level never occurs. To see why, consider the case where the firm has two managers IIs in period t. Then, these two individuals recruited in period $t-1$ as two workers Is must be promoted in the same period, a contradiction with our assumption that only one worker can be promoted in each period.

[11]Note that the cyclical overstaffing under Policy R-R results from the assumption of discrete number of entry-level workers. In a firm with a large number of workers, the law of large numbers always dictates a predictable set of winners of each age group, and thus the cyclicity of overstaffing disappears. In the working paper version (see https://fddi.fudan.edu.cn/fddien/8c/0f/c19513a297999/page.htm), we consider a large firm consists of a continuum of measure 2 of workers and promotes a continuum of measure 1 of workers in each period. In this situation, the firm is overstaffed at the managerial level in a steady way because there are 5/3 units of managers at all times. We show that the main conclusions of the model are robust to the change in the number of workers.

[12]A similar situation occurs when the monitoring quality of the firm in second best is such that it is almost perfect (i.e., $\Delta p \to 1/2$). In this sense, the firm is in a "quasi" first best in this section.

[13]Note that, given worker j selects $e_j = 0$, worker i will obtain W_i with probability $1/2 + \Delta p$ if selects $e_i = 1$, and obtain W_i with probability $1/2$ if selects $e_i = 0$. So worker i is willing to exert a high effort if and only if $(1/2 + \Delta p)W_i - \psi \geq W_i/2$ or $W_i \geq \psi/\Delta p$. Since $e_i = 1$ dominates $e_i = 0$ regardless of e_j, selecting $e_i = 1$ is a dominant strategy for worker i when $W_i \geq \psi/\Delta p$.

[14]See https://www.eeoc.gov/laws/types/age.cfm.

[15]See https://fddi.fudan.edu.cn/fddien/8c/0f/c19513a297999/page.htm.

[16]As another historical example, Popa (2015) shows that, in the 18th and early 19th centuries, sinecures were prevalent in British political system—the estimated number of major sinecures in the administration is one to two thousand. The most obvious sinecures were the ancient positions at the royal court, such as chamberlains, Grooms of the Royal Bedchamber, the Master of the Buckhounds, etc. Those who were appointed to these positions did not perform such menial jobs, but merely collected substantial payment from the government. Using data from the History of Parliament, he finds that such an appointee could earn a considerable income from about £ 500 a year to £ 2000 per year.

[17]See https://www.isa.org/standards-and-publications/isa-publications/intech-magazine/2008/may/final-say-top-heavy-management-drains-companies/.

[18]See https://www.forbes.com/sites/matthewherper/2013/09/19/merck-rd-head-bets-slashing-bureaucracy-will-unlock-innovation/#664ad94574fa.

[19]For public banks in Argentina, see Clarke and Cull (2002). For examples from Iran and China, see http://www.ft.com/cms/s/0/a048686a-9ed4-11e3-a48e-00144feab7de.html#axzz41mn735Ss and http://www.newschinamag.com/magazine/bloated-bureaucracy.

[20]See http://www.brookings.edu/research/opinions/2010/02/12-japan-economy-abe.

ACKNOWLEDGMENT

The authors thank seminar participants at the Norwegian School of Economics.

ORCID

Yanren Zhang https://orcid.org/0000-0001-5841-3666

REFERENCES

Acemoglu, D., Ticchi, D., & Vindigni, A. (2011). Emergence and persistence of inefficient states. *Journal of the European Economic Association*, *9*(2), 177–208.

Ahluwalia, M. S. (2002). Economic reforms in India since 1991: Has gradualism Worked? *Journal of Economic Perspectives*, *16*(3), 67–88.

Baker, G., Gibbs, M., & Holmstrom, B. (1994a). The wage policy of a firm. *Quarterly Journal of Economics*, *109*(4), 921–955.

Baker, G., Gibbs, M., & Holmstrom, B. (1994b). The internal economics of the firm: Evidence from personnel data. *Quarterly Journal of Economics*, *109*(4), 881–919.

Baker, G., Jensen, M. C., & Murphy, K. J. (1988). Compensation and incentives: Practice and theory. *Journal of Finance*, *43*(3), 593–616.

Banerji, A., & Sabot, R. H. (1993). *Wage distortions, overmanning, and reform in developing country public enterprises*. World Bank, Vice Presidency for Finance and Private Sector Development.

Becker, G. S. (1957). *The economics of discrimination* (2nd ed.). The University of Chicago Press.

Bhattacharya, S., & Guasch, J. L. (1988). Heterogeneity, tournaments, and hierarchies. *Journal of Political Economy*, *96*, 867–881.

Boudreau, K. J., Lakhani, K. R., & Menietti, M. (2016). Performance responses to competition across skill levels in rank-order tournaments: Field evidence and implications for tournament design. *Rand Journal of Economics*, *47*(1), 140–165.

Carmichael, H. L. (1983). The agents-agents problem: Payment by relative output. *Journal of Labor Economics*, *1*(1), 50–60.

Clarke, G. R., & Cull, R. (2002). Political and economic determinants of the likelihood of privatizing Argentine public banks. *Journal of Law and Economics*, *45*(1), 165–197.

Cremer, J. (1986). Cooperation in ongoing organizations. *Quarterly Journal of Economics*, *101*, 33–49.

Darity, W., & Mason, P. A. (1998). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives*, *12*, 63–90.

DeVaro, J. (2006). Internal promotion competitions in firms. *RAND Journal of Economics*, *37*(3), 521–542.

Fairburn, J. A., & Malcomson, J. M. (2001). Performance, promotion, and the Peter principle. *Review of Economic Studies*, *61*(1), 45–66.

Gibbon, E. (1797). *The history of the decline and fall of the Roman empire* (Volume the First, A New Edition). Printed for A. Strahan; and T. Cadell Jun. and W. Davies (Successors to Mr. Cadell) in the Strand, M.DCC.XCVII.

Goltsman, M., & Mukherjee, A. (2011). Interim performance feedback in multistage tournaments: The optimality of partial disclosure. *Journal of Labor Economics*, *29*(2), 229–265.

Grimshaw, D., Beynon, H., Rubery, J., & Ward, K. (2002). The restructuring of career paths in large service sector organizations: 'Delayering', upskilling and polarisation. *Sociological Review*, *50*(1), 27–115.

Hvide, H. K. (2002). Tournament rewards and risk taking. *Journal of Labor Economics*, *20*(4), 877–898.

Imhoff, L., & Kräkel, M. (2016). Ex post unbalanced tournaments. *Rand Journal of Economics*, *47*(1), 73–98.

Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, *76*, 323–329.

Kahn, C., & Huberman, G. (1988). Two-sided uncertainty and up-or-out contracts. *Journal of Labor Economics*, *6*, 423–444.

Kandori, M. (1992). Repeated games played by overlapping generations of players. *Review of Economic Studies*, *59*, 81–92.

Ke, R., Li, J., & Powell, M. (2018). Managing careers in organizations. *Journal of Labor Economics*, *36*(1), 197–252.
Lazear, E. P. (1991). Labor economics and the psychology of organizations. *Journal of Economic Perspectives*, *5*(2), 89–110.
Lazear, E. P. (1999). Personnel economics: Past lessons and future directions. *Journal of Labor Economics*, *17*(2), 199–236.
Lazear, E. P., & Oyer, P. (2012). Personnel economics. In R. Gibbons, & J. Roberts (Eds.), *The handbook of organizational economics* (pp. 479–519). Princeton University Press.
Lazear, E. P., & Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, *89*(5), 841–864.
Littler, C. R., Wiesner, R., & Dunford, R. (2003). The dynamics of delayering: Changing management structures in three countries. *Journal of Management Studies*, *40*(2), 225–256.
Malcomson, J. M. (1984). Work incentives, hierarchy, and internal labor markets. *Journal of Political Economy*, *92*(3), 486–507.
McDonald, l. M., & Solow, R. (1981). Wage bargaining and employment. *American Economic Review*, *71*, 896–908.
McDonald, l. M., & Solow, R. (1985). Wages and employment in segmented labor market. *Quarterly Journal of Economics*, *100*(4), 1115–1141.
McKee, M., & Wintrobe, R. (1993). The decline of organizations and the rise of administrators: Parkinson's law in theory and practice. *Journal of Public Economics*, *51*, 309–327.
Medoff, J. L., & Abraham, K. G. (1980). Experience, performance, and earnings. *Quarterly Journal of Economics*, *95*(4), 703–736.
Neumark, D., & Stock, W. A. (1999). Age discrimination laws and labor market Efficiency. *Journal of Political Economy*, *107*, 1081–1125.
Niskanen, W. N. (1968). The peculiar economics of bureaucracy. *American Economic Review*, *58*(2), 293–305.
Oswald, A. J. (1985). The economic theory of trade unions: An introductory survey. *Scandinavian Journal of Economics*, *87*(2), 160–193.
Parkinson, N. C. (1957). *Parkinson's law and other studies in administration*. Houghton Mifflin Company.
Popa, M. (2015). Elites and corruption: A theory of endogenous reform and a test using British data. *World Politics*, *67*(2), 313–352.
Roett, R. (1999). *Brazil: Politics in a patrimonial society*. Praeger.
Ross, S. L. (2003). *What is Known about testing for discrimination: Lessons learned by comparing across different market* (working paper). University of Connecticut, Storrs, Connecticut, USA. https://opencommons.uconn.edu/econ_wpapers/200321/
Schotter, A., & Weigelt, K. (1992). Asymmetric tournaments, equal opportunity laws and affirmative action: Some experimental results. *Quarterly Journal of Economics*, *106*, 513–539.
Shapiro, C., & Stiglitz, J. E. (1984). Equilibrium unemployment as a worker discipline device. *American Economic Review*, *74*, 433–444.
Tsoulouhas, T. (1999). Do tournaments solve the two-sided moral hazard problem? *Journal of Economic Behavior and Organization*, *40*, 275–294.
Van Biesebroeck, J. (2003). Technology choice: An application to automobile assembly. *Review of Economic Studies*, *70*, 167–198.
Waldman, M. (2012). Theory and evidence in internal labor markets. In R. Gibbons, & J. Roberts (Eds.), *The handbook of organizational economics* (pp. 520–574). Princeton University Press.
Williamson, O. E. (1964). *The economics of discretionary behavior: Managerial objectives in a theory of the firm*. Carnegie Institute of Technology.
Yinger, J. (1998). Evidence on discrimination in consumer markets. *Journal of Economic Perspectives*, *12*(2), 23–40.
Zabojnik, J. (2012). Promotion tournaments in market equilibrium. *Economic Theory*, *51*(1), 213–240.
Zeng, Q., & Yang, Y. (2017). Informal networks as safety nets: The role of personal ties in China's anti-corruption campaign. *China: An International Journal*, *15*(3), 26–57.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hvide, H. K., & Zhang, Y. (2021). Too big to succeed? Overstaffing in firms. *J Econ Manage Strat*. 30, 784–798. https://doi.org/10.1111/jems.12439