Rat Races and Glass Ceilings

Peter Bardsley  Katerina Sherstyuk
University of Melbourne*  University of Hawaii

March 2006

Abstract

In an ongoing organization, such as a large law partnership firm, employees are motivated not only by current rewards but also by the prospect of promotion, and the opportunity to make the rules in the future. This leads to a recursive contract design problem in an overlapping generations environment, where current agents may become future principals. The principal offers, and promotion-motivated agents accept, harsh rat race contracts with low wages and high effort levels. Hiring and promotion probabilities emerge as the preferred instrument to screen high cost workers, who face employment barriers and a glass ceiling.

Keywords recursive contracts, mechanism design, overlapping generations, adverse selection.

JEL D82,J41

*Corresponding author. Department of Economics, University of Melbourne, Victoria 3010 Australia, email p.bardsley@unimelb.edu.au.
1 Introduction

In an ongoing firm, employees are motivated not only by current rewards but also by the prospect of promotion. Promotion to a senior position brings with it the valuable opportunity to influence policy, and to make the rules in the future. In particular, it may bring the right to set the terms under which the next cohort of employees will work.

This process may be observed to some degree in many large organizations, but it is particularly clear in firms that are organized as professional partnerships; for example law firms, accountancy partnerships and management consultancy firms. In these firms the career path leads towards ownership rights, and the ability to make management decisions unrestricted by external owners. In these firms the prospect of future partnership may be far more important than the current wage in compensating for effort. In such a firm, the employment contract is recursive. The contract that will be agreed today depends upon the prospects for promotion and the nature of the contracts that will be agreed tomorrow. This structure leads to a dynamic programming problem in contract design. In this paper we consider such recursive contracts both in a perfect information environment and under adverse selection.

We find that the recursive structure already leads to inefficiently high effort levels and low wages — a rat race. Under adverse selection, an additional effect emerges. Employees entering the firm have different private characteristics that affect the cost of effort. Output is perfectly observable in our model; what is not observed is the private effort needed to produce that output. The principal will seek to extract information rents from the low (that is to say, high cost) types, in order to induce effort from the high (low cost) types and to discourage them from imitating the low types. Since wages are already low and effort levels high, due to the rat race, hiring and promotion probabilities emerge as the preferred screening instrument. We show that adverse selection leads to a downward bias in promotion probabilities for low types. We call this a glass ceiling.

As an application of these ideas, we use this framework to discuss possible implications for employees’ career paths in a dynamic contracting environment under equal opportunity laws and conventions. Some employees (typically women) may have a higher outside value of time due to child care costs
and household production opportunities\textsuperscript{1}, or some social groups may have suffered educational disadvantage. Suppose such characteristics are either unobservable, or they are observable, but firms cannot contract on them due to explicit legal requirements or to the force of social norms and conventions. This leads to an adverse selection problem. As we have seen, under this adverse selection pressure the recursive structure of the ongoing firm leads not only to a rat race, but also to entry and promotion barriers, or a glass ceiling. We thus find that the issues of high work levels (the “rat race”) and of unequal promotion rates (the “glass ceiling”) may be intimately interconnected. Curiously, we find that the adverse selection setting is welfare enhancing (at least relative to the very severe rat race that would ensue under full information). However the benefit accrues entirely to the top type, typically men. The presence of the bottom type mitigates the severity of the rat race, and the top type is able to earn some information rents.

We do not assert that this provides a comprehensive explanation of all manifestation of the rat race or the glass ceiling, which are clearly very complex phenomena; we do however put forward an environment in which these phenomena emerge very naturally as the result of rational contract design, and in which some intriguing connections between them become apparent\textsuperscript{2}.

There is an extensive literature on dynamic mechanism design, considered as the efficient design of incentive mechanisms in dynamic environments (see for example, Doepke and Townsend \cite{12}, Fernandes and Phelan \cite{15} and Marcet and Marimon \cite{30}). In this literature a single timeless principal designs, once and for all, a mechanism to influence the behavior of agents who live in a dynamic environment. The agents in such a model have to solve a recursive dynamic programming problem, and the principal in designing the mechanism has to take into account this dynamic programming problem. Our framework is quite different. We are concerned with an overlapping generations environment in which the mechanism is itself dynamic, being redesigned every period. Every period a new principal arises from among the

\textsuperscript{1}Wood, Corcoran and Courant \cite{39} provide evidence of the high private costs incurred by women in a large law firm.

\textsuperscript{2}The large commercial law firm, as we note below, seems in some respects to be a striking real life embodiment of this environment. One of the authors of this paper is married to a senior partner in a very large law firm. A significant motivation for this paper has been to understand the high effort levels, the career structures, the incentive structures, and the compensation disparities between partners and associates that have been evident from casual observation of this and similar firms.
agents of the previous period and has the option to redesign the contract under which the next generation of agents will work. The mechanism is itself determined recursively.

Landers, Rebitzer and Taylor [26] and Ghatak, Morelli and Sjostrom [19] both study principal-agent problems in overlapping generations environments. However their models are somewhat different from ours. Landers, Rebitzer and Taylor [26] look at adverse selection in law firms in a competitive environment with a restricted range of design instruments. We consider a full range of contract design instruments. We shall discuss the relationship between their work and ours in more detail below. Ghatak, Morelli and Sjostrom [19] show that, in the presence of borrowing constraints, workers may work harder today to save capital to become entrepreneurs tomorrow. In their model, such dynamic incentives are welfare-improving since they help to alleviate a moral hazard problem. In contrast, in our model there is no moral hazard and dynamic incentives (promotion prospects) severely reduce the agent welfare and lead to a rat-race.

Rat race phenomena have been studied extensively. Akerlof [1] made the original contribution, introducing an adverse selection or hidden type model to explain the existence of inefficiently high effort levels. Workers know their own type, and are sorted by a self selection mechanism (assembly lines that run at different speeds). This model was formalized in a competitive environment by Miyazaki [31] and Stiglitz [38]. More recent contributions are reviewed in Sampson and Simmons [36]; see also Holmstrom [23]. Rat races can also be generated in moral hazard environments, for example by tournaments that are designed to reduce shirking (see the survey by Gibbons and Murphy [20] and, in the context of legal firms, Ferrall [16]). In our model, rat races emerge not due to the moral hazard or adverse selection effects, but due to the recursive nature of the contract per se. That is to say, they already emerge in a complete information setting, where moral hazard or adverse selection effects are absent.

Apparently the term glass ceiling was popularized in a 1986 Wall Street Journal article. It has since generated a large popular and empirical liter-

\[ \text{See the discussion in Section 3.} \]

\[ \text{Further, in Ghatak et al. [19] the agents’ borrowing constraints create the dynamic incentive in the first place; in contrast, dynamic incentive in our model is present due to the recursive nature of the contract itself. We do assume that agents cannot borrow money. However, in our case this wage non-negativity constraint puts a limits on the rat race, rather than makes agents work harder; see section 3 below.} \]
ature. There is extensive evidence that women and minorities are underrepresented at the senior management levels of large firms (see Federal Glass Ceiling Commission [13]; Athey, Avery and Zemsky [2]; and Spurr [37] for the legal profession in particular). However the nature of the promotion barriers that they face is less clear (see also Blau and Kahn [7], Booth et al [8], Bulow and Summers [9], Groot et al [22], Lazear and Rosen [29], Wood et al [39], and Winter-Ebner and Zweimuller [40]). Our contribution is to show how differential promotion may emerge naturally as the preferred compensation and screening instrument in overlapping generation firms, and to point out that rat race and glass ceiling phenomena may be interlinked.

Many authors have discussed the role of internal promotions and up-or-out contracts in partnerships and other organizations (e.g., Landers, Rebitzer and Taylor [26]; Ferrall [16]; Khan and Huberman [24]; Prendergast [35]; Morrison and Wilhelm [32]; see also Baker, Jensen and Murphy [3]). The existing literature studies the role of internal promotions to either screen for better partners in an adverse selection setting (e.g., Landers, Rebitzer and Taylor [26]); or to control for agents’ or both principal’s and agents’ moral hazard problem (Ferrall [16]; Khan and Huberman [24]; Prendergast [35]); or to maintain the high value of the firm in human capital-intensive production settings (Morrison and Wilhelm [32]). In this paper we advance a different explanation of why promotion probabilities may emerge as a preferred compensation instrument in a dynamic contract setting. The promise of tomorrow’s promotion to partnership is used by the current principal to extract high efforts from employees today, to the extent that leads to an over-work, or a rat-race. Since compensation with the promise of the firm’s tomorrow’s profit is costless for the current principal (who is short-lived and only cares about today’s profits), the partnership will always use future promotions as a preferred compensation instrument. Moreover, given that the future reward is sufficiently high (and the agents are not too risk-averse), probabilistic promotions can replace certain ones, and, ex-ante, can be used to compensate all workers, even if only a few will be actually promoted. We do not explicitly model how uncertainty about the promotions is realized ex-post. What

---

5 Some authors argue that tournaments or up-or-out contracts are an effective way to resolve uncertainty about promotions when the principal cannot otherwise commit to rewarding agents’ good outcomes, i.e., in a dual moral hazard setting (e.g., Khan and Huberman [24]; Prendergast [35], Bhattacharya and Guasch [5]). Others point out drawbacks of using tournaments (e.g., Lazear [28]). In our setting, the principal is short-lived and has no incentives not to reward the agents as promised.
matters is that employees enter the firm with an expectation of possibility of promotion, which differs across types.

The law firm is a natural laboratory in which to explore recursive contracts in an overlapping generations environment, and this is the example that we will have in mind. Rat races and glass ceilings appear to be very prevalent in such firms, which are typically organized as an ongoing partnership with no outside equity. However we believe that similar effects will operate in many organizations where internal promotion is an important part of the career structure. The characteristics of the large commercial law firm have been well documented in a number of studies, for example Carr and Matthewson [10], Feinberg [14], Ferrall [16], Galanter and Palay [18], Gilson and Mnookin [21], Landers Rebitzer and Taylor [26], O'Flaherty and Siow [34], and Spurr [37]. In the traditional law partnership the partners own the firm in equal shares and they face soft incentives (for example equal profit shares and permanent tenure). Associates in these firms face much harsher incentives, with an up or out career path. They earn a relatively low wage (compared with a partner’s profit share), and they are subject to work pressures that appear to be inefficiently high (Ferrall [16], Landers, Rebitzer and Taylor [26]; see also Akerlof [1]). Traditionally, major law firms were an exclusively male preserve. The proportion of women associates grew rapidly from this low base line in the 1980’s, and stabilized quite quickly at levels representative of the general population (Galanter and Palay [18]); however a glass ceiling still seems to operate, and far fewer reach partnership (Spurr [37], Wood, Corcoran and Courant [39]). The issues of excessively high effort levels and of barriers to the promotion of women remain very relevant in these firms (Groot et al [22], Blau and Kahn [7], Landers, Rebitzer and Taylor [26], Spurr [37]).

Several recent studies model the law firm as an overlapping generations structure, but not from a full contract design perspective (e.g., Ferrall [16]; see also Cremer [11]). In particular, Landers, Rebitzer and Taylor [26] present an interesting overlapping generations adverse selection model of the law firm. It may be useful to point out how their work differs from what we present here. Their model is driven by free riding under equal sharing between partners, and high effort levels are imposed at the entry level to screen for good quality potential partners. In our model there is a single principal\textsuperscript{6}. We focus

\textsuperscript{6}Or a multi-partner coalition of principals who have no conflict of interest and act like a single principal; their only task is contract design, on which they all agree, and they
on the dynamic inefficiencies that can arise under recursive contracting, and on mechanisms that screen agents, not principals. They describe equilibrium outcomes under a particular given contractual structure. We allow the principal to design the contractual structure optimally. They assume that promotion probabilities are fixed, and screen only through wages. We allow both promotion probabilities and wages to be used as screening instruments; in contrast to Landers et al. we find that wages are in fact never used. As a consequence, our model is sufficiently rich that we are able to explore not only effort (rat race) effects but also promotion (glass ceiling) effects and their interaction. It is interesting that both models, although so different in structure and assumptions, predict inefficiently high effort levels.

The structure of the paper is as follows. The basic model is set out in Section 2. The optimal recursive contract is derived in Section 3, under both complete (3.2) and incomplete information (3.3), and a numerical example is described in 3.4. The implications of the model are discussed in Section 4.

2 The Model

There are $n_i$ agents of type $\theta_i$ born each period, and each agent lives for two periods. The agent’s type is always known to the agent himself, and may be either observable to the principal (as we assume in Section 3.2), or be the agent’s private information. The type space $\{\theta_i : i = 1, \ldots, I\}$ is finite, and $\theta_I$ is the highest type. The set of types and the number of agents of each type is common knowledge; the total number of agents born each year is $n = \sum_i n_i$. The agent’s type affects his cost of effort, as will be defined below. Utility is separable between periods and there is no discounting; all agents are risk neutral.\textsuperscript{7} Agents are born with no initial wealth and are credit constrained; this implies that agents’ wages will be constrained to be non-negative in our model.\textsuperscript{8}

\textsuperscript{7}These assumptions are for the sake of simplicity. Qualitatively similar results may be obtained in a model with moderately risk averse-agents, and with moderate discounting; see Bardsley [4].

\textsuperscript{8}This assumption is consistent with the absence of entry fees and bonds in actual employee contracts; see Baker, Jensen and Murphy [3] for discussion.

exert no effort. The assumption that the principal has no productive role may appear counter-intuitive. We impose this assumption to demonstrate that even if an agent’s ability does not matter once she becomes a principal, ability may still play an important role in determining promotion rates.
There is a single firm, owned in each period by a single principal (we explicitly exclude the possibility of any outside ownership\textsuperscript{9}). The firm is assumed to exist indefinitely into the past and into the future, so there will be no initial or terminal effects.\textsuperscript{10} Similar to Landers, Rebitzer and Taylor [26], we assume that complete contracting on output is possible. We allow the firm to decide how many agents of each type $\theta_i$ to employ at date $t$ (the hiring probability $\sigma_{i,t}$), what they are to do (the required output level $b_{i,t}$), how much they are paid (the wage $w_{i,t}$), and who will be promoted (the promotion probability $\pi_{i,t}$).\textsuperscript{11} To simplify notation, we write $w_i = w_{i,t}$ for the current period wage, $w_i^+ = w_{i,t+1}$ for the next period’s wage. A similar notational convention will be used for other variables.

Unemployed agents, and agents who reject the contract, receive a reservation utility of 0 in both periods. Agents who are not promoted leave the firm and receive 0 in the second period. The principal exerts no effort. Her only task is to design the contract to be offered to the next generation. In order to continue the law firm analogy, we may from time to time refer to the principal as the partnership\textsuperscript{12}.

An agent of type $\theta_i$ who produces output $b \in \mathbb{R}_+$ exerts effort $e_i = e(b, \theta_i)$. Each agent’s output is ex-post observable, but the effort is not. We make the following assumptions about the technology and the population.

**Assumption 0 (maximum output)** There is a (possibly type dependent) maximum feasible output $b^*(\theta)$, and $e(b, \theta) \rightarrow \infty$ as $b \rightarrow b^*(\theta)$.

**Assumption 1 (convexity of effort)** The effort function is continuously

\textsuperscript{9}This may be for regulatory reasons in the case of law firms. For the widespread existence of the closely held firm see La Porta et al [27] and Morrison and Wilhelm [32].

\textsuperscript{10}Of course no real firm, or real economy, exists for ever. The infinite horizon OLG model may be considered to be an approximation to the long lived firm in a steady state where any initial or terminal effects are remote and unimportant.

\textsuperscript{11}See the discussion in Section 1 on probabilistic promotions.

\textsuperscript{12}It is straightforward to re-interpret this as a model of a multi-principal partnership firm under the assumption that the partners exert no effort, and that they share the partnership profits equally (for details, see Bardsley [4]). Although they may be of different types, under these assumptions the multiple principals have no difficulty in agreeing on the objectives of the firm. We will not explore this interpretation any further here, and henceforth we will retain the assumption that the number of partners is 1.
differentiable at least twice, and

\[ e(0, \theta) = 0 \]  \hspace{1cm} (1a)
\[ e_b(b, \theta) > 0 \]  \hspace{1cm} (1b)
\[ e_{bb}(b, \theta) > 0 \]  \hspace{1cm} (1c)

for \( 0 \leq b < b^*(\theta) \).

**Assumption 2 (sorting conditions)** If \( \theta_i < \theta_j \) and \( 0 \leq b < b^*(\theta_i) \), then

\[ e_b(b, \theta_i) > e_b(b, \theta_j) \]  \hspace{1cm} (2a)
\[ e(b, \theta_i) - e(b, \theta_j) \to \infty \text{ as } b \to b^*(\theta_i) \]  \hspace{1cm} (2b)
\[ e_{bb}(b, \theta_i) \geq e_{bb}(b, \theta_j) \]  \hspace{1cm} (2c)

**Assumption 3 (productive types)** All types are potentially productive, where we say that a type \( \theta \) is potentially productive if \( e(b, \theta) < b \) for some \( 0 \leq b < b^*(\theta) \), where \( b^*(\theta) \) is defined in Assumption 0 above.

**Assumption 4 (no rare types)** \( n_i \geq 1 \) for all \( i \).\textsuperscript{13}

All assumptions are standard (Fudenberg and Tirole [17], Laffont and Martimort [25]). Assumption 0 is introduced for technical convenience and will be used to show the existence of a stationary optimal contract. Assumption (2a) is the usual single crossing condition, which says that marginal effort is declining in type, while (2c) says that the convexity of the effort curve (which may be interpreted as aversion to output uncertainty) is also declining in type. This will guarantee that the virtual effort curve (to be defined below) is convex, and will allow us to show that the optimal employment rule is deterministic, and the optimal contract monotone.

Given the assumptions of risk-neutrality and no discounting between periods, a type \( \theta_i \) agent’s expected utility, conditional on being employed, is his wage net of the cost of effort, plus the expected payoff \( W^+ \) from becoming the principal in the next period:

\[ U_i = w_i + \pi_i W^+ - e(b_i, \theta_i). \]

\textsuperscript{13}This assumption may seem to be unnecessary under the single principal interpretation that we emphasise. If however we consider a normalised model of a multi-principal firm as mentioned in the previous footnote, then this assumption has some bite and needs to be made explicit. Under this interpretation, \( n_i \) is the expected number of agents of type \( \theta_i \) per principal (the “type specific span of control”).
His ex ante expected utility, prior to any employment contract being offered, is $\sigma_i U_i$.

Since the principal is not directly involved in production, her objective is simply to maximize her own current payoff $W$, which is the firm’s total output less the wages paid to the agents. Assuming the agents’ types are unobservable to the principal, an optimal contract maximizes the payoff to the principal subject to ex ante individual rationality, incentive compatibility and wage non-negativity constraints. By the revelation principle (e.g., Myerson [33]), the principal can use direct revelation mechanisms to reveal the agents’ types. The design problem is thus as follows.

**Problem (P)** Choose $w_i, \sigma_i, \pi_i, b_i$, to maximize

$$W = \sum_i n_i \sigma_i (b_i - w_i)$$  \hspace{1cm} (3a)

subject to

$$\sigma_i \left( w_i + \pi_i W^+ - e(b_i, \theta_i) \right) \geq 0$$  \hspace{1cm} (3b)
$$\sigma_i \left( w_i + \pi_i W^+ - e(b_i, \theta_i) \right) \geq \sigma_j \left( w_j + \pi_j W^+ - e(b_j, \theta_i) \right)$$  \hspace{1cm} (3c)
$$b_i \geq 0$$  \hspace{1cm} (3d)
$$w_i \geq 0$$  \hspace{1cm} (3e)
$$0 \leq \pi_i \leq 1$$  \hspace{1cm} (3f)
$$0 \leq \sigma_i \leq 1$$  \hspace{1cm} (3g)
$$\sum_i n_i \pi_i \sigma_i = 1.$$  \hspace{1cm} (3h)

Here (3b) and (3c) are individual rationality and incentive compatibility constraints, and (3d-3h) are feasibility constraints. If we let

$$\gamma_{ij}(b) = e(b, \theta_i) - e(b, \theta_j)$$

be the cost advantage of type $j$ over type $i$ in producing output $b$, and eliminate wages $w_i$ in favour of rents $U_i$, we can rewrite the system as follows.

**Problem (P')** Choose $U_i, \sigma_i, \pi_i, b_i$, to maximize

$$W = W^+ + \sum_i n_i \sigma_i (b_i - e(b_i, \theta_i) - U_i)$$  \hspace{1cm} (4a)
subject to

\[
\begin{align*}
\sigma_i U_i & \geq 0 \\ 
\sigma_i \gamma_{i,j}(b_i) & \leq \sigma_j U_j - \sigma_i U_i \leq \sigma_j \gamma_{i,j}(b_j) \\ 
b_i & \geq 0 \\ 
U_i + e(b_i, \theta_i) & \geq \pi_i W^+ \\ 
0 & \leq \pi_i \leq 1 \\ 
0 & \leq \sigma_i \leq 1 \\ 
\sum_{i} n_i \pi_i \sigma_i & = 1.
\end{align*}
\]

The incentive compatibility constraints (4c) are here written in a form such that they are required to hold only for \(i < j\). The sorting and convexity assumptions guarantee that if \(i < j\) then \(\gamma_{i,j}(b)\) is positive, strictly increasing, convex in \(b\), and \(\gamma_{i,j}(b) \to \infty\) as \(b \to \infty\).

3 The Optimal Contract

The problem \((P')\) is a standard adverse selection problem except for two features. First, it is recursive. To solve for the current period value \(W\) one must know the future value, \(W^+\), of being principal in the next period. To solve for that one must know \(W^{++}\), and so on . . . . So we have a dynamic programming problem in contract design. Second, due to the fixed firm size assumption, we have extra feasibility constraints 4f and 4h, which limit the total amount of promotions that the principal can give out to the agents to unity. This imposes a certain “budget balance in promotion probabilities” requirement which is atypical for standard adverse selection problems, and which complicates the solution. We deal with the dynamic aspect of the problem first. Then, in Section 3.1, we suggest a way to determine promotion probabilities in the optimal contract, once other features of the contract are characterized. Finally, we solve for the optimal contracts in Sections 3.2 and 3.3.

As with any dynamic programming problem, there are two situations where the solution is straightforward. One is the finite horizon problem, when the firm ceases to exist at some future date \(t = T\). We can then set \(W^{(T+1)} = 0\), and solve by backwards induction. The other is to look for
stationary steady state solutions, closing the system by setting $W^+ = W$. In what follows we will focus on the steady state optimal contract\textsuperscript{14}.

The assumption of agent rationality is sufficient to determine the system dynamics (the transition from $W$ to $W^+$). However there will be many possible dynamic paths, distinguished by different terminal conditions or transversality conditions. We assume that agents coordinate their expectations on one such path. Specifically, we assume that agents expect that $W^+ = W$, where $W$ is a value consistent with a stationary equilibrium. We ensure the existence of such a $W$ in the following Lemma.

Lemma 1 (existence) There exists a stationary optimal contract. Furthermore, in this contract $W = W^+ > 0$.

Proof. See Appendix A.

We note that the non-negativity constraint on wages is essential for this existence result. Without a lower bound on wages no equilibrium exists. The principal will demand an arbitrarily large participation payment from the agent. The risk neutral agent is happy to comply, since this cost may be passed on to the next generation. The contract thus degenerates into a Ponzi scheme.\textsuperscript{15}

\textsuperscript{14}It is not hard to show that the finite horizon problem will converge to a steady state as $T \to \infty$. Bardsley [4] discusses this under the assumptions of perfect information and homogeneous agents.

\textsuperscript{15}Formally, suppose there is no wage constraint of the type 3e, $w_i \geq w_i$, for each $\theta_i$; that is, wages are not bounded from below. Suppose an optimal stationary contract exists and let $(w_i, \sigma_i, \pi_i, b_i)$, for all $\theta_i$, be such a contract, with the corresponding level of principal’s profit $W$. By stationarity, $W^+ = W$. Take an $\varepsilon > 0$, and consider an alternative contract which pays each type less by $\pi_i \varepsilon$ in wages than the original contract, with all other elements of the original contract unchanged: $(\tilde{w}_i, \tilde{\sigma}_i, \tilde{\pi}_i, \tilde{b}_i) = (w_i - \pi_i \varepsilon, \sigma_i, \pi_i, b_i)$ for all $\theta_i$. From 3a, the new contract is preferred by the principal: $\tilde{W} = \sum_i n_i \sigma_i (b_i - w_i + \pi_i \varepsilon) = W + \varepsilon > W$. Since the wage does not enter the constraints 3d, 3f-3h, we only need to check the IR and IC constraints 3b and 3c to establish the feasibility on the new contract. By stationarity, we have $\tilde{W}^+ = \tilde{W}$, and hence $\tilde{\sigma}_i \left( \tilde{w}_i + \tilde{\pi}_i \tilde{W}^+ - \varepsilon \left( \tilde{b}_i, \theta_i \right) \right) = \sigma_i \left( w_i - \pi_i \varepsilon + \pi_i W^+ + \pi_i \varepsilon - \varepsilon \left( b_i, \theta_i \right) \right) = \sigma_i \left( w_i + \pi_i W^+ - \varepsilon \left( b_i, \theta_i \right) \right) \geq 0$, which establishes 3b. Similarly, to establish that the new contract satisfies the IC constraint 3c requires to show that $\sigma_i \left( w_i - \pi_i \varepsilon + \pi_i W^+ + \pi_i \varepsilon - \varepsilon \left( b_i, \theta_i \right) \right) \geq \sigma_j \left( w_j - \pi_j \varepsilon + \pi_j W^+ + \pi_j \varepsilon - \varepsilon \left( b_j, \theta_i \right) \right)$, which holds as long as the IC constraints hold for the original contract. We thus found another feasible contract that is strictly preferred by the principal, a contradiction.

This reasoning also shows that the wage constraint 3e binds for all types in an optimal
3.1 The Relaxed Problem

We will now show, under assumptions that will hold in the steady state, that we can replace problem $P_0$ with simpler relaxed problems ($P_{00}$, under complete information, in Section 3.2, and $P_{000}$, under incomplete information, in Section 3.3).

Let us for the moment put aside the question of how the parameters $\sigma_i, U_i, b_i$ might be determined, and consider the promotion probabilities $\pi_i$ that might implement contract $P'$. We note that the parameters $\pi_i$ do not occur directly in the principal’s objective (4a), and that they occur only in the constraints (4e, 4f, 4h). Constraint (4e) is the wage non-negativity constraint. Aggregating this constraint over all types gives us a global wage constraint

$$\sum_i n_i \sigma_i (U_i + e(b_i, \theta_i)) \geq W^+,$$

which just states that the aggregate wage bill is non-negative. We note that this also is free of the parameters $\pi_i$.

Now the principal does not care directly about the parameters $\pi_i$. She cares only that there exist feasible parameters that can implement her desired contract. The following Lemma shows when such feasible parameters exist.

Lemma 2 Assume that the parameters $0 \leq \sigma_i \leq 1, U_i \geq 0, b_i \geq 0$ are fixed, and assume that $W^+ > 0$. Then

1. There exist parameters $\pi_i \geq 0$ satisfying (4e) and (4h) if and only if the global wage constraint (5) is satisfied;

2. The parameters $\pi_i$ are unique for all employed types (that is, types such that $\sigma_i > 0$) if the global wage constraint (5) binds;

3. If, in addition, employment is deterministic (that is, $\sigma_i = 0$ or $\sigma_i = 1$ for all $i$) then (4f) is satisfied as well.

Proof. See Appendix A. ■

stationary complete information contract (when the IC constraint 3c is absent because the types are observable and contractible upon), and that it binds for at least one employed type in the optimal incomplete information contract. We will further show that, in fact, the wage constraints will bind for all types under both information structures.
Corollary 1 If $W^+ > 0$ and employment is deterministic then we may relax Problem (P') by deleting the parameters $\pi_i$ and all the constraints (4e, 4f, 4h) that depend on these parameters, and replacing them by the single global wage constraint (5).

Corollary 2 If the global wage constraint (5) binds then the wage and promotion parameters implied by the relaxed problem are

\[
\begin{align*}
  w_i &= 0 \\
  \pi_i &= \frac{U_i + e(b_i, \theta_i)}{W^+}
\end{align*}
\]

for all employed types. Without loss of generality we may assume that these formulas hold also for unemployed types with $\sigma_i = 0$.

The analogues of Lemma 2 and Corollaries 1, 2 hold as well under full information, omitting the incentive compatibility constraints (4c).

Let us consider, informally, what it means for the wage constraint (5) to bind. As discussed above, the problem (P') differs from the standard agency problem in that the principal has, in addition to the wage, another means to reward the agents. This is $W^+$, the value of inheriting the firm. This prize, which is of no intrinsic benefit to the principal\(^{16}\), becomes valuable if she can use it to extract a greater surplus from the agents.

In any agency model the principal would like, on the one hand, to maximize the surplus earned by the firm and, on the other, to extract this surplus from the agents. In the standard one-shot model, at least in a complete information environment, there is no conflict between these objectives. The principal can maximize the surplus by requiring production efficiency and can extract all of this surplus by reducing wages until all agents are at their participation margin. Under incomplete information this is no longer possible, and effort will be biased downward (except for the top type) below the efficient level, for reasons of incentive compatibility.

Here we have something different. If $W^+$ is sufficiently large, she may not be able to extract all of the agents’ surplus, even under complete information, just by reducing wages. If the wage constraint binds then the only instrument that she can use to extract rents is the effort level. So as $W^+$ becomes

\(^{16}\)By assumption she cannot sell the firm to an outsider. She cannot sell it to the agents since they are initially endowed with zero wealth.
large, we may expect to see an upward distortion in effort levels, opposite in
direction to the downward distortion required by incentive compatibility. In
fact she may be willing to accept some production inefficiency (a rat race) in
order to extract rents. We shall show below that, in the steady state, \( W^+ \) is
sufficiently large to induce a large upward distortion of this kind.

### 3.2 The Complete Information Recursive Contract

Under complete information, we can drop the incentive compatibility con-
straints \((4c)\). Using the strategy set out in Section 3.1 above, we study the
following relaxed problem\(^{17}\) in which we drop all constraints containing the
\( \pi_i \), replacing them by the global wage constraint \((6e)\).

**Relaxed Complete Information Problem** \((P'')\) *Choose \( \sigma_i, b_i, U_i \) to max-
imize*

\[
W = W^+ + \sum_{i} n_i \sigma_i (b_i - e(b_i, \theta_i) - U_i) \quad (6a)
\]

subject to

\[
\begin{align*}
U_i & \geq 0 \quad (6b) \\
b_i & \geq 0 \quad (6c) \\
0 & \leq \sigma_i \leq 1 \quad (6d) \\
W^+ & \leq \sum_{i} n_i \sigma_i (U_i + e(b_i, \theta_i)) \quad (6e)
\end{align*}
\]

We first solve the relaxed problem \((P'')\) for \( \sigma_i, U_i \) and \( b_i \), for an arbitrary
fixed value of \( W^+ \), and confirm that this gives us a deterministic contract. In
fact, under complete information, we find that there is full employment with
\( \sigma_i = 1 \) for all \( i \). We also establish that, as in the standard principal-agent
complete information setting, all agents are driven to participation margin,
\( U_i = 0 \). Then, imposing the steady state condition \( W = W^+ \), which (from
\((6a)\)) now takes the form

\[
\sum_{i} n_i (b_i - e(b_i, \theta_i)) = 0, \quad (7)
\]

\(^{17}\)Note that we can write \((6b)\) as shown because the inequality systems \( \sigma U \geq 0, 0 \leq
\sigma \leq 1 \), and \( U \geq 0, 0 \leq \sigma \leq 1 \) are equivalent.
we verify that the global constraint (6e) binds. We then use Lemma 2 and Corollaries 1 and 2 to interpret what this means for \( w_i \) and \( \pi_i \).

**Proposition 1 (complete information)** The complete information steady state contract is characterized by

- full employment: \( \sigma_i = 1 \) for all \( i \)
- zero wages: \( w_i = 0 \) for all \( i \)
- output monotonicity: \( b_i \) is strictly increasing in \( i \)
- full rent extraction from every type: \( U_i = 0 \) for all \( i \)
- equal effort at the margin: \( e_b(b_i, \theta_i) = e_b(b_j, \theta_j) \) for all \( i, j \)
- inefficiently high effort levels (a rat race): \( e_b(b_i, \theta_i) > 1 \) for all \( i \)
- dissipation of all surplus: \( \sum_i n_i (b_i - e(b_i, \theta_i)) = 0 \).

**Proof.** See Appendix A.

While full employment, full rent extraction, output monotonicity and equal marginal efforts across types are standard characteristics of optimal contracts under complete information, other features are not. Specifically, zero wages, inefficiently high effort levels, and dissipation of all surplus are the unique characteristics of the optimal contract that are driven by the contract’s recursive structure. The complete information contract for the case of two types is illustrated in Figure 1. The Proposition tells us that the weighted average of the contract points \( (b_1, e(b_1, \theta_1)) \) and \( (b_2, e(b_2, \theta_2)) \) lies on the 45° line (if \( n_1 = n_2 \) then the weighted average is the midpoint; it always lies on the interval between the two contract points). The figure also demonstrates that under complete information both types of agents are necessarily employed, with \( b_1 > 0, b_2 > 0 \).

We find that wages are driven to minimal levels. Obviously agents’ wages are never in reality actually zero, just low compared to the principal. The large gap between associates’ and partners’ wages in law firms is well recognized and documented. For example Landers Rebitzer and Taylor [26] report that the average annual salary of associates in their survey was $79,154 as compared to $247,067 for partners; the gap in top tier firms is even greater.
It is worth noting one result that we do not get. It is not in general true that promotion is monotonic in type. Equality of marginal effort does not imply monotonicity of total effort, nor of promotions (see Figure 2). There is no merit per se in this model in promoting high quality agents to the partnership, since partners perform no productive role. In contrast, as we will show below, promotions will become monotonic in type when the adverse selection effects are present (Section 3.3, Proposition 2).

The main conclusion that we draw from the complete information model is that the recursive contracting structure leads to an upward distortion in effort levels and a severe rat race. From Equation (6a), noting that $W = W^+$ and $U_i = 0$, we see that the principal will in fact increase effort levels until

$$\frac{\sum_i n_i e_i (b_i, \theta_i)}{\sum_i n_i b_i} = 1.$$

The average total effort equals 1, and the value of the output is completely dissipated by the effort of producing it.

It is easy to see that this outcome is inefficient. A contract is efficient if no generation of agents can be made better off without reducing the welfare of another generation, and no agent within a generation can be made better off without hurting someone else. Since the current principal’s utility is already accounted for as the utility of one of the agents in the previous generation, only the ex ante utilities of agents (potential employees) enter the efficiency criterion. Under the optimal stationary contract described above, all individual rationality constraints bind, and hence the ex ante utility of every generation of agents is zero. It is easy to show that there exists a feasible contract that yields to every generation a positive utility (consider, for example, the output maximizing contract in which $e_b (b_i, \theta_i) = 1$, $w_i = e (b_i, \theta_i)$, $\sigma_i = 1$ and $\pi_i = \frac{1}{n}$).

---

18 There may be good reasons to do otherwise, in order to induce high effort levels from high cost types. One can show that promotion probability is increasing in type provided that the effort function meets the following regularity assumption: $e_\theta (b, \theta) e_\theta (b, \theta) \geq e_b (b, \theta) e_\theta (b, \theta)$. This assumption is satisfied for many reasonable specifications of the effort function, including for example $e (b, \theta) = \frac{b^2}{\theta}$ and $e (b, \theta) = \frac{b}{\theta} + b^2$.

19 As noted above, the firm is assumed to exist indefinitely into the past and future, so there are no initial or terminal effects.
Figure 1: The complete information contract \((b^*, e^*)\) is a weighted average of points on the effort curves with equal marginal effort, and lies on the diagonal.
Figure 2: Non-monotonic promotion. Equal marginal effort does not imply effort (or promotion) monotonicity.
3.3 The Incomplete Information Recursive Contract

It is convenient at this point to define virtual effort functions\(^{20}\)

\[
\varepsilon(b, \theta_i) = \begin{cases} 
  e(b, \theta_i) + \frac{N_{i+1}}{n_i} \gamma_{i,i+1}(b), & \text{if } i < I \\
  e(b, \theta_I), & \text{if } i = I.
\end{cases}
\]

where \(I\) is the top type, \(\gamma_{ij}(b) = e(b, \theta_i) - e(b, \theta_j)\) is the cost advantage of type \(j\) over type \(i\) in producing output \(b\) (this was defined in Section 2), and \(N_{i+1} = n_{i+1} + \ldots + n_I\) is the number agents of a higher type than \(\theta_i\). We note that, by Assumptions 2a and 2c, the virtual effort \(\varepsilon(b, \theta_i)\) is increasing and strictly convex in \(b\), and it satisfies \(\varepsilon(0, \theta_i) = 0\).

If we follow the standard approach of dropping all individual rationality constraints except the first, and all but the adjacent downward incentive compatibility constraints, this leaves us with the following active incentive compatibility and rationality constraints:

\[
\sigma_1 U_1 = \sigma_1 \frac{n_1}{N_1} \zeta_1
\]  
\[
\sigma_{i+1} U_{i+1} = \sigma_i U_i + \sigma_i \gamma_{i,i+1}(b_i) + \sigma_{i+1} \frac{n_{i+1}}{N_{i+1}} \zeta_{i+1},
\]  

where we have introduced variables \(\zeta_i \geq 0\) to capture the active constraints. These variables represent the excess rents, above and beyond those required by individual rationality and incentive compatibility, earned by type \(\theta_i\). The scaling factors \(\frac{n_i}{N_i}\) are introduced for notational convenience. We notice that

\[
\sigma_j U_j = \sum_{i=1}^{j-1} \sigma_i \gamma_{i,i+1}(b_i) + \sum_{i=1}^{j} \sigma_i \frac{n_i}{N_i} \zeta_i
\]  

and hence that

\[
\sum_i n_i \sigma_i (U_i + e(b_i, \theta_i)) = \sum_i n_i \sigma_i (\zeta_i + \varepsilon(b_i, \theta_i)).
\]

We also note the monotonicity condition

\[
\sigma_i \gamma_{i,i+1}(b_i) \leq \sigma_{i+1} U_{i+1} - \sigma_i U_i \leq \sigma_{i+1} \gamma_{i,i+1}(b_{i+1})
\]  

\(^{20}\)The interpretation of the virtual effort is standard for adverse selection models; see, e.g., Fudenberg and Tirole [17], pp. 266 and 287.
which is implied by incentive compatibility.

Once again, using the strategy set out in Section 3.1, we replace $P$ by a relaxed problem $P''$ in which we drop all constraints containing the $\pi_i$, replacing them by the global wage constraint (12e). As discussed above, we also drop the individual rationality constraints and the incentive compatibility constraints except for (8, 9). Using these to eliminate the $U_i$ in favour of the $\zeta_i$, we arrive at the following \footnote{See Footnote 17 for the justification for using the constraint $\zeta \geq 0$, rather than $\sigma \zeta \geq 0.$}.

**Incomplete Information Problem ($P''$)** Choose $\sigma_i, b_i, \xi_i$ to maximize

$$W = W^+ + \sum_{i} n_i \sigma_i \left( b_i - \varepsilon (b_i, \theta_i) - \zeta_i \right)$$

subject to

$$\zeta_i \geq 0$$
$$b_i \geq 0$$
$$0 \leq \sigma_i \leq 1$$
$$W^+ \leq \sum_{i} n_i \sigma_i \left( \zeta_i + \varepsilon (b_i, \theta_i) \right)$$

The system $P''$ is essentially identical to the complete information system $P''$, with the $U_i$ replaced by the $\zeta_i$ and effort replaced by virtual effort. The only significant difference is that virtual effort need not satisfy the analogue of Assumption 3 (Productive Types).

There is a natural indeterminacy associated with contracts in which agents are either not employed ($\sigma_i = 0, b_i \geq 0, U_i \geq 0$) or employed but inactive ($\sigma_i \geq 0, b_i = 0, U_i = 0$). In the first the agent is not employed, but receives a contract contingent on the (non-existent) employment. In the second the agent is employed but idle and unrewarded. Let us define a null contract to be one such that $\sigma_i b_i = 0$. From the monotonicity condition (11) we see that if some type is offered a null contract then all lower types must also receive a null contract. From (8, 9, 11) we see that in any null contract we must have $\sigma_i U_i = \sigma_i \zeta_i = 0$. Clearly there are many null contracts $(\sigma_i, b_i, \zeta_i)$ with $\sigma_i b_i = \sigma_i \zeta_i = 0$ that are completely equivalent from the point of view of the agents and the principal. To remove this indeterminacy we will impose the convention that the only null contract that is offered is the zero contract $(\sigma_i = 0, b_i = 0, \zeta_i = 0)$.\footnote{See Footnote 17 for the justification for using the constraint $\zeta \geq 0$, rather than $\sigma \zeta \geq 0.$}
Proposition 2 (incomplete information) The incomplete information steady state contract is characterized by

- zero/one employment by type: $\sigma_i = 0$ or $\sigma_i = 1$ for all $i$

- sequential employment: $\sigma_i$ is non-decreasing in $i$

- zero wages: $w_i = 0$ for all $i$

- output monotonicity: $b_i$ is non-decreasing in $i$, and strictly increasing for participating types

- promotion monotonicity: $\pi_i$ is non-decreasing in $i$, and strictly increasing for participating types

- monotonic rents: $U_i$ is non-decreasing in $i$, and strictly increasing for participating types

- full rent extraction at the bottom: $U_1 = 0$

- equal virtual effort at the margin: $\varepsilon_b(b_i, \theta_i) = \varepsilon_b(b_j, \theta_j)$ for all participating types $i, j$

- inefficiently high effort levels (a rat race) at the top: $e_b(b_1, \theta_1) > 1$

- dissipation of the virtual surplus: $\sum_i n_i (b_i - \varepsilon(b_i, \theta_i)) = 0$.

Proof. See Appendix A

Some features of the optimal contract, such as output monotonicity, monotonic rents, full rent extraction at the bottom, and equal virtual effort at the margin, are typical for most adverse most selection problems. However, other features, such as zero wages, a rat race at the top, and dissipation of the virtual surplus, are unique and due to the recursive nature of the problem. We also note that, unlike the complete information case, promotions are now monotonic in type.

The optimal contract is illustrated in Figure 3 for the two type case. It is clear why there may be no stationary full employment contract that meets the first order conditions. If the virtual effort curve $\varepsilon(b_1, \theta_1)$ lies above the $45^\circ$ line, and if the contract points $(b_1, \varepsilon(b_1, \theta_1))$ and $(b_2, \varepsilon(b_2, \theta_2))$ are chosen to equalize marginal virtual cost, then it is possible that the weighted average
Figure 3: The incomplete information contract is a weighted average of points on the virtual effort curves with equal marginal virtual effort. No full employment equilibrium exists if such points lie always above the diagonal.
of the contract points will lie always above the 45° line, which would violate the stationarity condition.

Thus, similarly to a non-recursive contracting setting, we obtain that information effects may endogenously limit the size of the firm. High type agents are hired before low type agents, and low type agents may not be hired at all, even though they are potentially productive.

It is also clear from Figure 3 why \( e_b(b_2, \theta_2) > 1 \). For assume that \( e_b(b_2, \theta_2) \leq 1 \). If both types are employed then the contract point \((b_1, \varepsilon(b_1, \theta_1))\) must lie at a point on the virtual effort curve where \( \varepsilon_b(b_1, \theta_1) = \varepsilon_b(b_2, \theta_2) = e_b(b_2, \theta_2) \leq 1 \). But by convexity, marginal virtual effort exceeds average virtual effort, so \( \frac{\varepsilon(b_1, \theta_1)}{b_1} < 1 \). Thus the contract point must lie below the 45° line. This is impossible, by the same argument as in the complete information case. If only type \( \theta_2 \) is employed then the contract point \((b_2, \varepsilon(b_2, \theta_2))\) must lie where the effort curve intersects the 45° line, and it is clear then that \( e_b(b_2, \theta_2) = \varepsilon_b(b_2, \theta_2) > 1 \).

We see that, as in the complete information case, the recursive structure again leads to an upward bias in effort levels. In particular, the effort level of the top type is unambiguously above the efficient production level. However this upward bias is counteracted by the downward bias introduced by the requirements of incentive compatibility. Under complete information the average effort is 1, but (by 12a, noting that \( W^+ = W \) and \( \zeta_i = 0 \))

\[
\frac{\sum_i n_i \varepsilon_i(b_i, \theta_i)}{\sum_i n_i b_i} = \frac{\sum_i n_i \varepsilon(b_i, \theta_i)}{\sum_i n_i b_i} = 1.
\]

Here we are summing only over employed types; the inequality is strict if at least two types are employed. So the average effort is reduced by adverse selection, provided that at least two types are employed.

Once again, the wage and the effort level are the preferred instruments for extracting rent from the agents, and wages are driven to zero. Both output and promotion are now monotonic in type. High types produce more, and they are promoted more often to guarantee that they earn positive rents. Since there are no productivity reasons to promote low types less often (see Section 3.2 for characteristics of the complete information contract for comparison), these differences in promotion rates are due to adverse selection. We will interpret this below as a glass ceiling effect.

We note that welfare is higher under incomplete information than under complete information. Now some agents earn positive information rents;
under complete information all rents are dissipated. Thus the incomplete information is welfare improving, at least relative to the inefficiency of the complete information contract. The presence of low types (provided that they are employed) mitigates the severity of the rat race, and higher types are better off. Obviously, downward distortion of output and the presence of incomplete information rents for higher type agents are standard result for adverse selection models. However, while in the standard case these distortions reduce efficiency, in our case they increase the dynamic efficiency of the contract.

### 3.4 An Example

We have seen that in an overlapping generations firm with recursive contracts the principal will impose a “rat-race” contract. Wages are driven down to the reservation level of zero. Effort is driven up to inefficiently high levels; if there are no information constraints effort will be driven up to the point where the surplus is entirely dissipated. As in most principal-agent models, all rent is extracted from the bottom. Since wages are already zero and effort is high, this can only be done by reducing the promotion probability of the bottom type (a “glass ceiling”). Thus “rat-race” and “glass ceiling” effects are intimately interconnected. In order to sustain very high effort levels at the top, the bottom type is driven to the participation margin by low promotion rates.

We illustrate these features of the optimal contract with a numerical example. Let the effort functions be

\[ e(b, \theta_1) = \tau b + b^2 \]
\[ e(b, \theta_2) = \tau b \]

where \(0 < \tau < 1\). Let \( \mu = \frac{n_1}{n_1 + n_2} \) be the proportion of type \( \theta_1 \) agents in the population.

---

22The effort functions in this example do not satisfy Assumption 0 (finite output ceiling). However this assumption is used only to show the existence of an optimal contract, which can be done directly here. Alternatively, these effort functions could be modified by adding a smooth penalty function that cuts in at some very high effort level; this would not affect the optimal contract as calculated.
It is straightforward to calculate that if $\mu < \frac{\tau}{2}$ then only the top type is employed ($\sigma_1 = 0, \sigma_2 = 1$), and the top type receives a pure rat-race contract that extracts all rent and dissipates all surplus: $b_2 = 1$, $e_b(b_2, \theta_2) = 2$ and $e_b(b_2, \theta_2) = 1$. If $\mu \geq \frac{\tau}{2}$ then there is full employment ($\sigma_1 = \sigma_2 = 1$), with output levels

$$b_1 = \frac{\mu + \sqrt{\mu (\mu - 2\mu\tau + \tau^2)}}{2\mu} - \frac{\tau}{2\mu}$$

$$b_2 = \frac{\mu + \sqrt{\mu (\mu - 2\mu\tau + \tau^2)}}{2\mu}.$$

The properties of this example are illustrated in Tables 1 to 3, for $\tau = 0.6$. Table 1 shows how output, average effort and marginal effort vary as the population mix changes. Table 2 shows how the population mix of types is reflected in the employed population mix and in the population mix of partners (since we have normalized the number of partners to 1, this should

<table>
<thead>
<tr>
<th>Population Mix (%)</th>
<th>Output</th>
<th>Average Effort</th>
<th>Marginal Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type $\theta_1$</td>
<td>Type $\theta_2$</td>
<td>$b_1$</td>
<td>$b_2$</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td>0.17</td>
<td>0.92</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.26</td>
<td>0.86</td>
</tr>
<tr>
<td>60</td>
<td>40</td>
<td>0.32</td>
<td>0.82</td>
</tr>
<tr>
<td>70</td>
<td>30</td>
<td>0.35</td>
<td>0.78</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>0.38</td>
<td>0.75</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>0.39</td>
<td>0.72</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.40</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 1: Output and Effort
Table 2: Employment Mix and Expected Partnership Mix.

<table>
<thead>
<tr>
<th>Population mix (%)</th>
<th>Employment mix (%)</th>
<th>Partnership mix (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type $\theta_1$</td>
<td>Type $\theta_2$</td>
<td>Type $\theta_1$</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>80</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>60</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>70</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>80</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: Employment Mix and Expected Partnership Mix.
<table>
<thead>
<tr>
<th>Population Mix (%)</th>
<th>Promotion Probability</th>
<th>Information Rent</th>
<th>Payoff to principal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type $\theta_1$</td>
<td>Type $\theta_2$</td>
<td>Type $\theta_1$</td>
</tr>
<tr>
<td>10 90</td>
<td>-</td>
<td>$\frac{1}{n}$</td>
<td>-</td>
</tr>
<tr>
<td>20 80</td>
<td>-</td>
<td>$\frac{1}{n}$</td>
<td>-</td>
</tr>
<tr>
<td>30 70</td>
<td>0</td>
<td>$\frac{1.43}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>40 60</td>
<td>$\frac{21}{n}$</td>
<td>$\frac{1.53}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>50 50</td>
<td>$\frac{40}{n}$</td>
<td>$\frac{1.60}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>60 40</td>
<td>$\frac{56}{n}$</td>
<td>$\frac{1.66}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>70 30</td>
<td>$\frac{70}{n}$</td>
<td>$\frac{1.71}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>80 20</td>
<td>$\frac{81}{n}$</td>
<td>$\frac{1.75}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>90 10</td>
<td>$\frac{91}{n}$</td>
<td>$\frac{1.79}{n}$</td>
<td>0.00</td>
</tr>
<tr>
<td>100 0</td>
<td>$\frac{1}{n}$</td>
<td>-</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3: Promotion Probabilities and Expected Payoffs. (Recall that $n$ is the size of the population of potential employees.)
be interpreted as the expected number of partners of each type). Table 3 shows how the promotion probabilities, the information rents earned by each type, and payoff to the principal vary.

If the proportion of type $\theta_1$ agents in the population is below the critical level of 30%, then the bottom type is completely excluded and only the top type is employed. These agents are offered a pure rat-race contract with a zero wage, and an effort level so high (the marginal effort of 2 is well above the efficient level of 1) that the surplus is totally exhausted (the average effort is 1). As a consequence, all ex ante rents are dissipated.

As the proportion of type $\theta_1$ agents rises a threshold is crossed, and it suddenly becomes profitable to employ all of the type $\theta_1$ agents as well as all of the type $\theta_2$ agents. If even one type $\theta_1$ agent is employed it is necessary to pay information rents to all of the type $\theta_2$ agents. Below the threshold there are not enough potential type $\theta_1$ employees to warrant doing this. Above the threshold, once the first type $\theta_1$ agent is employed then these rents must be paid in full, so there is no reason not to employ them all.

As the proportion of type $\theta_1$ agents rises beyond this critical level the population mix is immediately reflected in the employed population, since everybody is employed. However the number who make it through into partnership is quite small. For example, if the population is evenly balanced, with 50% of each type, only 20% of the partnership is of type $\theta_1$ while 80% are of type $\theta_2$. By way of comparison, under complete information 47% would be of type $\theta_1$ and 53% of type $\theta_2$. Even if the proportion of type $\theta_2$ agents fell to 10%, they would remain grossly over-represented in the partnership at around 18% (under complete information this would only be 12%). We interpret this as a glass ceiling effect.

4 Discussion

We believe that this model throws light on questions of over-work and differentiated career paths in professional partnership firms. We show that future career concerns and the possibility of becoming a partner in the future may result in severe over-work (rat race) and low wages for the agents. While several earlier papers also show that dynamic incentives make agents work harder (Holmstrom [23]; Ghatak et al. [19]), our insight is quite different. Holmstrom [23] shows that a rat race may be caused by the agent’s desire to signal their high ability, when wages reflect employer’s expectation of agent
ability. In Ghatak et al. [19], workers motivated by future certain rewards work harder than they would in a static setting; in their setting the dynamic aspect improves the welfare, as it alleviates the moral hazard problem. In contrast, we suggest an explanation for a rat-race phenomenon that is unrelated to either moral hazard, or agents’ desire to signal their ability. Even though our model assumes heterogeneous agents (which further allows us to deal with glass ceiling effect), it is straightforward to show that the rat-race result would carry to a homogeneous agents setting as well. We demonstrate, under both complete and information setting, that in long-lived organizations with “patient” risk-neutral agents, promotion becomes principal’s preferred compensation instrument, leading to low wages and over-work for the agents. Although it may appear that the model is driven by the assumptions of agent risk-neutrality and no future discounting, qualitatively similar results may be obtained for moderately risk averse-agents, and with moderate discounting; see Bardsley [4].

We also show that in an adverse selection setting, promotions may be used as the preferred screening instrument, thus leading to the glass ceiling phenomenon. This suggests an explanation for differentiated career paths for different types of workers in professional partnerships. Consider a complete information environment where there are two observable types. One is of lower productivity within the firm. For example, women may have higher expected effort costs due to child-bearing and child-rearing conventions; or minority groups may have language problems or they may have faced educational discrimination. For clarity of exposition we will refer to the two types as “women” and “men.” In a complete information environment we would expect to see full employment with discriminatory contracts.

If types are unobservable, or they are observable, but no contract can be conditioned on type, either by law or by social convention, then these effectively become hidden types. We then see some interesting possibilities. In the first place, there may be segregated firms which employ only men, even though there are potentially productive women in the population. So there is a threshold effect. Until the number of women in the population (in our example this would be the number of women law graduates entering

\[1, \text{ pp. 603-4}\]

\[23\]
the employment market) reaches a critical level, none are hired. Once the
threshold is crossed, they are hired but they face a promotion barrier. There
is an obvious coordination problem that may discourage any women from
investing in legal qualifications.

Secondly, even if both types are employed we expect to see systematic
differences in promotion rates, the differences that cannot be attributed solely
to differences in ability. Very few women are promoted to the top of the
organization. It should be emphasized that this is not a result of explicit
discrimination. The apparently discriminatory outcomes are the result of
self selection induced by the structure of the contracts.

In the mixed firm the presence of women mitigates the high rat-race effort
levels. As a consequence, men now earn positive information rents. Thus in
this model the hidden type setting is welfare improving for the agents (relative
to the highly inefficient complete information rat-race, which extracts all
rents from everyone). The women are no worse off, while men are better off.
It is curious that it is the top type that benefits from non-discriminatory
hiring and promotion.

The existence of glass ceiling effects has been extensively documented
(see Federal Glass Ceiling Commission [13], and other references cited in
Section 1). Models of this effect provide no convincing explanation of why
promotion emerges as the preferred screening instrument; for example Bulow
and Summers [9] and Lazear and Rosen [29] assume that it is expensive to
train and promote women because they are more likely to leave. Athey
et al [2] assume that mentoring occurs predominately within rather than
between types. These models appear to be applicable across the whole job
spectrum (and have been tested predominately against such data; see for
example Booth et al [8]), whereas the broad evidence is that glass ceiling
effects are most pronounced at the senior executive level. Our model is
focused in particular on promotion to this level. We find that promotion is
preferred to wages as a screening instrument because the current principal
cares only about her return, not the total profits of the firm. Any attempt to
screen through wage dispersion would lower her payoff, but screening through
promotions is, to her, costless. Our model also shows that there can be an
interaction between rat race and glass ceiling effects.

It is possible that there is more than one effect operating. The mechanism that we
describe applies to promotion to the senior executive level, not to promotions lower down
within the firm.
We suggest that the model that we have presented is consistent with three of the main stylized facts which characterize the large commercial law firm (see Galanter and Palfay [18] and the other references cited in Section 1):

- very high, and apparently inefficient, effort levels
- the rather rapid historical transition of these firms from male only institutions to institutions with a balanced intake of men and women at the entry level
- the stubborn glass ceiling effect that seems to prevent a balanced proportion of women from reaching partnership.

Of course there are other effects that we have ignored. Principal among these are all the issues connected with shirking and moral hazard (see Ferrall [16]), screening effects designed to improve the quality of the partnership (see Landers, Rebitzer and Taylor [26]), and competition between firms (see Bernhardt and Scoones [6]). It would be interesting, but beyond the scope of this model, to explore the relative importance and interaction of these effects.
A Appendix

Proof of Lemma 1 (existence). Let us first solve the principal’s problem, and show that an optimal contract \((U_i, \pi_i, \sigma_i, b_i)\) exists, conditional on the parameter \(W^+ \geq 0\). We will then show that for an appropriate value of \(W^+\) the contract is stationary. Without loss of generality, we may assume that \(\sigma_i > 0\) for all \(i\). For if not, just exclude all types for which \(\sigma_i = 0\). There is a finite number of subsets of the type space that we might wish to exclude in this way; if we can show existence of an optimal contract in each of these cases then then we need only to search over this finite number of cases to find a contract that is optimal over all.

We note that there is an individually rational, incentive compatible contract that yields a strictly positive surplus \(W\) to the principal. By the productive types assumption 3 there is a feasible output level \(b_1\) such that \(b_1 > e(b_1, \theta_1)\). By the single crossing assumption we have in fact that \(b_1 > e(b_1, \theta_1) > e(b_1, \theta_i)\) for all \(i > 1\). Consider the pooling contract in which all agents are employed, all are asked to produce output \(\tilde{b} = b_1\), all are paid a wage \(\tilde{w} = e(b_1, \theta_1) < b_1\), and all are promoted with equal probability \(\tilde{p} = \frac{1}{n}\). An agent of type \(\theta_i\) gets utility \(\tilde{w} + \tilde{W} + e(b_1, \theta_i)\) \(\tilde{W}\), and \(\tilde{W} \geq 0\) so this individually rational; by construction it offers the principal a strictly positive surplus. This example also shows that the set of feasible contracts is non-empty. We can thus assume, without loss of generality, that

\[
W = W^+ + \sum_i n_i \sigma_i (b_i - e(b_i, \theta_i) - U_i) > 0. \tag{13}
\]

Let us now hold the parameters \(\sigma_i\) fixed and solve the problem conditional on the value of these parameters. First we show that the \(U_i\) are uniformly bounded. The function \(b - e(b, \theta_i)\) is continuous and concave on \([0, b^* (\theta_i)]\), and it tends to \(-\infty\) as \(b \to b^* (\theta_i)\). Thus it is bounded above, and this bound may be taken to be uniform in type (we have only a finite number of types). The function \(\sum_i n_i \sigma_i (b_i - e(b_i, \theta_i))\) is thus bounded above uniformly in the \(b_i\). Thus \(\sum_i n_i \sigma_i U_i\) is bounded above. Since the terms in this sum are non-negative, the quantities \(U_i\) are uniformly bounded.

We now bound the effort levels. By the sorting condition (1b), if \(i < I\) then the incentive compatibility condition (4c) must eventually be violated as \(b_i \to \infty\). Thus we can bound \(b_i\) by a constant \(B_i < b^* (\theta_i)\). The constraints (4c, 4e) bound \(b_I\) only from below, and as \(b_I\) becomes large it drives the objective function negative. The point at which this happens may depend
on the parameters $U_i, \pi_i, (1 \leq i \leq I)$ and $b_i (1 \leq i < I)$, but these are now bounded, so we may without harm impose a bound $b_I \leq B_I < b^* (\theta_I)$. All the variables $(U_i, \pi_i, b_i)$ are now uniformly bounded so we can appeal to the Weierstrass Maximum Theorem to show that there exists an optimal contract.

Thus for any value of the parameters $0 \leq \sigma_i \leq 1$ there is an optimal contract with value $W(W^+, \sigma_1, \ldots, \sigma_I)$. By the Berge Maximum Theorem this value is continuous in the $\sigma_i$, so, appealing once again to Weierstrass, there there exists an optimal value for these parameters. Thus we have shown that there exists an optimal contract, contingent on $W^+$, with value $W(W^+)$. We now consider the question of stationarity. Appealing again to the Berge Maximum Theorem, we note that $W(W^+)$ is a continuous function. We noted above that the principal can always guarantee a strictly positive surplus provided that $W^+ \geq 0$. Thus $W(W^+) > W^+$ for $W^+ = 0$. But we also know the $W$ is bounded above. The easiest way to see this is from equation $(3a)$, since the $b_i$ are bounded by the limits $b^* (\theta_i)$. Thus $W(W^+) < W^+$ for $W^+$ very large. By the Intermediate Value Theorem there exists a value $W^+$ such that $W(W^+) = W^+$. This gives us the required stationary contract.

Proof of Lemma 2. Since (5) is the aggregation of the constraints (4e) it is clearly a necessary condition. Conversely, let $\tilde{\pi}_i = \frac{U_i + e(b_i, \theta_i)}{W^+}$. Constraint (4e) can be written $\pi_i \leq \tilde{\pi}_i$, and (5) becomes $\sum_i n_i \tilde{\pi}_i \sigma_i \geq 1$. If this constraint binds, then it is clear that $\pi_i = \tilde{\pi}_i$ will do the job, and that (except for unemployed types with $\sigma_i = 0$) these are the only parameters that will do so. If the constraint is slack, then there is more than one way to find a feasible set of the $\pi_i$. For example, we can set $\pi_i = \lambda \tilde{\pi}_i$, where $\lambda$ is a suitably chosen parameter.

To show that (4f) is satisfied we must confirm that $\pi_i \leq 1$ if $\sigma_i = 1$. Assume not, and let $i_0$ be such that $\pi_{i_0} > 1$ and $\sigma_{i_0} = 1$. Then, by the “no rare types” Assumption 4,

$$1 = \sum_i n_i \tilde{\pi}_i \sigma_i \geq n_{i_0} \pi_{i_0} \sigma_{i_0} > n_{i_0} \geq 1.$$

Proof of Proposition 1 (complete information). The Lagrangean
The first order and complementary slackness conditions are

\[ \alpha_i = (1 - \lambda) \sigma_i \quad \text{(14a)} \]
\[ b_i + \delta_i (1 - 2\sigma_i) = (1 - \lambda) (e (b_i, \theta_i) + U_i) \quad \text{(14b)} \]
\[ \sigma_i + \beta_i = (1 - \lambda) \sigma_i e_b (b_i, \theta_i) \quad \text{(14c)} \]
\[ \lambda \perp \sum_i n_i \sigma_i (U_i + e (b_i, \theta_i)) - W^+ \quad \text{(14d)} \]
\[ \alpha_i \perp U_i \quad \text{(14e)} \]
\[ \beta_i \perp b_i \quad \text{(14f)} \]
\[ \delta_i \perp \sigma_i (1 - \sigma_i) \quad \text{(14g)} \]

We note first that in an optimal contract we cannot have \( \sigma_i = 0 \) or \( b_i = 0 \) for any type. For if so, we could find a new contract in which agents of this type were employed and produced a strictly positive surplus (by Assumption 3 all types are potentially productive). Such agents could be paid a wage to compensated them for the increased effort, and offered a share of the surplus to induce them to participate, but offered no additional possibility of promotion. Such a contract would benefit the principal, and would not interfere with the contracts offered to other types.

Thus we may assume that \( \sigma_i > 0 \) and \( b_i > 0 \) (and hence \( \beta_i = 0 \)). By (14c) we see that \( \lambda < 1 \), so marginal effort is equalized across types. By the single crossing condition, this implies that output is monotonically increasing in type. By (14a) we see that \( \alpha_i > 0 \), and hence that \( U_i = 0 \).

Let us now assume that \( 0 < \sigma < 1 \), so that \( \delta_i = 0 \). Then, combining (14b) and (14c) we see that \( e_b (b_i, \theta_i) = \frac{\rho(e_i)}{\rho_i} \). Marginal effort equals average effort. But this is not possible; effort is strictly convex and \( b_i > 0 \), so marginal effort is strictly greater than average effort. Thus \( \delta_i > 0 \), and hence \( \sigma_i = 1 \).
We now focus on the steady state contract, and show that $\lambda > 0$. Assume the contrary. Then by (14c) the marginal effort $e_b(b_i, \theta_i) = 1$ for each type. The contract thus maximizes the surplus $b_i - e(b_i, \theta_i)$, type by type. By Assumption 3 (productive types) this maximized surplus will be strictly positive. Thus $b_i > e(b_i, \theta_i)$; but aggregating over types produces a contradiction to the steady state condition (7). Thus $\lambda > 0$.

It then follows that the global constraint (6e) binds. Then, by Lemma 2 and Corollaries 1, 2, there exists a unique set of feasible payments $w_i$ and promotion probabilities $0 \leq \pi_i \leq 1$ that implement the optimal contract, and furthermore, $w_i = 0$ for all $i$. Finally, we also see that $e_b(b_i, \theta_i) = \frac{1}{1-\lambda} > 1$, so effort levels are higher than is consistent with efficient production.

**Proof of Proposition 2 (incomplete information).** We will solve the relaxed problem $P''$. It will then be necessary to check that the resulting contract is deterministic ($\sigma_i = 0$ or $\sigma_i = 1$), that the global wage constraint (12e) binds, and that the omitted individual rationality and incentive compatibility constraints hold.

The Lagrangian is

$$L = W^+ + \sum_i n_i \sigma_i (b_i - \varepsilon(b_i, \theta_i) - \zeta_i)$$

$$\quad + \lambda \left( \sum_i n_i \sigma_i (\zeta_i + \varepsilon(b_i, \theta_i)) - W^+ \right)$$

$$\quad + \sum_i n_i \alpha_i \zeta_i + \sum_i n_i \beta_i b_i + \sum_i n_i \delta_i \sigma_i (1 - \sigma_i)$$

The first order and complementary slackness conditions are

$$\alpha_i = (1 - \lambda) \sigma_i$$

$$b_i + \delta_i (1 - 2\sigma_i) = (1 - \lambda) (\varepsilon(b_i, \theta_i) + \zeta_i)$$

$$\sigma_i + \beta_i = (1 - \lambda) \sigma_i e_b(b_i, \theta_i)$$

$$\lambda \perp \sum_i n_i \sigma_i (\zeta_i + \varepsilon(b_i, \theta_i)) - W^+$$

$$\alpha_i \perp \zeta_i$$

$$\beta_i \perp b_i$$

$$\delta_i \perp \sigma_i (1 - \sigma_i)$$

We show first that $\sigma_i = 0$ or $\sigma_i = 1$. Assume that $\sigma_i > 0$. So we are not dealing with a null contract. Thus $b_i > 0$ and $\beta_i = 0$. Then, by (15c),
\( \lambda < 1 \). By (15a) we see that \( \alpha_i > 0 \) and \( \zeta_i = 0 \). If \( 0 < \sigma_i < 1 \) then \( \delta_i = 0 \), so from (15b) and (15c) we would have \( \varepsilon_b (b_i, \theta_i) = \varepsilon(b_i, \theta_i) \); but this is impossible since \( \varepsilon (b_i, \theta_i) \) is strictly convex in \( b_i \), \( \varepsilon (0, \theta_i) = 0 \), and \( b_i > 0 \). Thus \( \sigma_i = 0 \) or \( \sigma_i = 1 \). We have also shown that \( \sigma_i \zeta_i = 0 \) (this is trivial if \( \sigma_i = 0 \)), so the bottom individual rationality and the downward adjacent incentive compatibility constraints bind. We also note that \( \delta_i > 0 \), at least for all employed types; by Lemma 1 (Existence) there must be at least one employed type.

From the monotonicity condition (11), if \( \sigma_i = 1 \) then \( \sigma_{i+1} = 1 \) (since, by the null contracts convention, \( b_i > 0 \)). This shows that employment is monotonic. Given that \( \sigma_i = 1 \) for all participating types, the monotonicity condition also implies that output is strictly increasing across participating types.

We now consider rents. If types \( \theta_i, \theta_{i+1} \) participate, then by equation (10), \( U_{i+1} - U_i = \gamma_i (b_i) > 0 \); so rents are monotonic.

We now introduce the steady state condition \( W = W^+ \) which here takes the form

\[
\sum_i n_i b_i = \sum_i n_i \varepsilon (b_i, \theta_i) .
\] (16)

We show that \( \lambda > 0 \). Assume the contrary. Then, by (15b), \( b_i \geq \varepsilon (b_i, \theta_i) \) since \( \delta_i \geq 0 \), with strict inequality for some type. But aggregating this relationship yields a contradiction to (16). Thus the global constraint (12c) binds. Then, by Lemma 2, there exists a unique set of feasible payments \( w_i \) and promotion probabilities \( \pi_i \), with \( 0 \leq \pi_i \leq 1 \), which implement the optimal contract, and furthermore, \( w_i = 0 \) for all \( i \). We also see that \( \varepsilon (b_i, \theta_i) = \varepsilon (b_i, \theta_i) = \frac{1}{1-\lambda} > 1 \), so effort levels are inefficiently high at the top.

Let us consider the question of monotonic promotion. By equation (10) we see that

\[
\sigma_i \pi_i W = \sigma_i U_i + \sigma_i \varepsilon (b_i, \theta_i) = \sum_{j=1}^{i-1} \sigma_j \gamma_{j,j+1} (b_j) + \sigma_i \varepsilon (b_i, \theta_i)
\]

since \( \sigma_i \zeta_i = 0 \). Assume that types \( \theta_i, \theta_{i+1} \) participate. Then

\[
\pi_{i+1} W - \pi_i W = \gamma_{i,i+1} (b_i) + \varepsilon (b_{i+1}, \theta_{i+1}) - \varepsilon (b_i, \theta_i)
\]

\[= \varepsilon (b_{i+1}, \theta_{i+1}) - \varepsilon (b_i, \theta_{i+1}) > 0 \]
since output is monotonic.

Finally, we show that the omitted constraints (4c) and (4d) hold. First of all, by (9), expected rent \( \sigma_i U_i \) is monotonic. Thus the higher individual rationality constraints must hold. We consider the upward incentive compatibility constraints. Let \( i < j \); then

\[
\sigma_j U_j - \sigma_i U_i = \sum_{k=i}^{j-1} \sigma_k \gamma_{k,k+1} (b_k)
\]

\[
\leq \sigma_j \sum_{k=i}^{j-1} \gamma_{k,k+1} (b_k)
\]

\[
\leq \sigma_j \sum_{k=i}^{j-1} \gamma_{k,k+1} (b_j)
\]

\[
= \sigma_j \gamma_{i,j} (b_j).
\]

The first line is justified by (9), the second by the monotonicity of \( \sigma_i \), the third by the monotonicity of \( b_i \) and the single crossing condition. The proof for the downward constraint is similar. ■
References


