

# Incentives in Markets, Firms, and Governments

Daron Acemoglu\*  
Massachusetts Institute of Technology

Michael Kremer\*\*  
Harvard University

Atif Mian\*\*\*  
University of Chicago

We construct a simple career concerns model where high-powered incentives can distort the composition of effort by inducing excessive signaling. We show that in the presence of this type of career concerns, markets typically fail to limit competitive pressures and cannot commit to the desirable low-powered incentives. Firms may be able to weaken incentives and improve efficiency by obscuring information about individual workers' contribution to output, and thus reducing their willingness to signal through a moral-hazard-in-teams reasoning. However, firms themselves have a commitment problem, since firm owners would like to provide high-powered incentives to their employees to increase profits. When firms cannot refrain from doing so, government provision may be useful as a credible commitment to low-powered incentives. Governments may be able to achieve this even when operated by a self-interested politician. Among other reasons, this may happen because of the government's ability to limit yardstick competition and reelection uncertainty. We discuss possible applications of our theory to pervasive government involvement in predominantly private goods such as education and management of pension funds.

## 1. Introduction

Although a range of transactions take place in markets and are subject to high-powered incentives, many important activities are organized within firms that are partly shielded from market incentives. Still others are conducted by

---

\*Department of Economics, Massachusetts Institute of Technology. Email: [daron@mit.edu](mailto:daron@mit.edu).

\*\*Department of Economics, Harvard University. Email: [mkremer@fas.harvard.edu](mailto:mkremer@fas.harvard.edu).

\*\*\*Graduate School of Business, University of Chicago. Email: [atif@chicagogsb.edu](mailto:atif@chicagogsb.edu).

We thank Avinash Dixit, Mathias Dewatripont, Robert Gibbons, Joseph Harrington, Bengt Holmstrom, Canice Prendergast, Raghuram Rajan, Jean Tirole, Jan Zabojsnik, two anonymous referees, and seminar participants at the National Bureau of Economic Research Economics Organization Conference for useful comments and Daniel Carvalho, Alexandre Debs, Jean Lee, Janina Matuszeski, and Courtney Umberger for research assistance.

governments, where they are even more insulated from market incentives. Although the costs of low-powered incentives are well known, a body of work beginning with Holmstrom and Milgrom (1991) suggests that in some circumstances low-powered (weak) incentives may actually be optimal. In this article, we suggest that in activities where low-powered incentives are optimal, governments may be the desirable form of organization because of their ability to *credibly* commit to such incentives. The model thus offers a new incentive-based explanation for why, despite their well-known inefficiencies, governments often provide private goods such as education, health care, and pensions.

In our model, workers with career concerns choose two types of effort, one which is socially productive and one which is socially unproductive, but affects observed performance. We illustrate the argument with the example of education, and assume that teachers make separate decisions about how much effort to exert in building children's underlying human capital and about how much effort to exert in "teaching to the test." High-powered incentives therefore have costs as well as benefits: they induce more productive effort but also more unproductive effort. If the distortionary effects of high-powered incentives are sufficiently severe, low-powered incentives may be optimal. However, competitive pressures in markets make incentives naturally high powered. Firms, on the other hand, may be able to "coarsify" information by organizing activity (e.g., teaching) into teams. Coarser information, in turn, creates a standard moral-hazard-in-teams problem and induces low-powered incentives. Nevertheless, since firms compete in the market and are the residual claimant of the profits they generate, firm owners face exactly the same strong incentives to improve their firms' observed performance. If they can secretly reward individual team members (teachers) on the basis of their performance, they will not be able to credibly commit to low-powered incentives, and the same inefficiencies present with markets will reemerge.

When both markets and firms fail to credibly commit to low-powered incentives, government operation may be a possible solution. We discuss a number of reasons within the framework of our model (and beyond those already emphasized in the literature, see, e.g., Dixit [1997, 2002]) about why governments can better commit to low-powered incentives. First, in the presence of common shocks even if the ability and actions of the politician matter and politicians are driven by self-interest, they may be able to commit to low-powered incentives because of the weakening of yardstick competition associated with government provision. Second, politicians may be less subject to career concerns incentives in the context of improving educational performance because of reelection uncertainty caused by other factors.

Our analysis therefore offers a new incentive-based explanation for why activities such as education and pension funds, where the true quality of output is not well observed and hence the risk of distortion toward the "bad" type of effort is particularly high, may be organized within governments.<sup>1</sup> This adds to

---

1. Wilson (1989) and Dixit (2002) emphasize the importance of "multitasking" issues in the public sector.

both public good and political economy theories of government. The former cannot explain why provision of private goods such as education, health care, and pensions accounts for a much larger fraction of government expenditure than public goods, such as national defense, scientific research, and interstate highways. Similarly, existing theories of government based on rent seeking, see, for example, Niskanen (1971), Bates (1981) and Shleifer and Vishny (1994), suggest that governments may be too large, but they do not explain why governments more often engage in operating hospitals, than, say, growing wheat or manufacturing pasta.

We should note from the outset, however, that this article does not offer a complete theory of the division of economic activities between markets, firms, and governments. Many other factors clearly influence the boundaries of firms (see, among others, Williamson [1985], Grossman and Hart [1986], and Hart and Moore [1990]), and our analysis abstracts from the most important determinants of government behavior, political economy concerns. Nevertheless, our model is complementary to existing theories of firms and governments, since it provides a simple unified framework for thinking about markets, firms, and governments. Although existing studies on the boundaries of the firm focus on how asset ownership shapes investment incentives, our approach is based on the ability of an organizational structure to credibly commit to manipulating information and shows how these factors may be important in determining whether some activities should be operated by firms or by governments.<sup>2</sup>

In addition, our article is most closely related to the career concerns literature (e.g., Stein 1989; Meyer and Vickers 1997; Dewatripont et al. 1999; Holmstrom 1999), and to the multitasking literature (e.g., Holmstrom and Milgrom 1991, 1994), which also emphasizes the costs of high-powered incentives.<sup>3</sup> Our model combines elements from both models and this combination is essential for our study of governments: government operation or regulation is useful precisely because the underlying career concerns problems make it impossible for the firms to commit to not rewarding employee success. In addition, the role of firms as institutions for suppression of information has been discussed by Gibbons and Murphy (1992), Gibbons (1998), and Acemoglu (1998), but not in a context where suppression of information is useful for weakening incentives and improving the composition of effort. Moreover, this work assumes that firms have no commitment problem and therefore provides no role for the government. Articles by Kremer (1997) and Levin and Tadelis (2002)

---

2. It is also worth emphasizing that our theory does *not* imply that government operate only those activities where the costs of weak incentives are outweighed by their benefits. Governments may grow beyond their “optimal” size because of a range of reasons, including those related to political economy and corruption, but still do so while specializing more in areas where they have a comparative advantage. Therefore, our theory is informative about the areas in which we may expect to see government operation, even when such behavior is not motivated by welfare-maximizing objectives.

3. The literature on advertising with imperfect information about quality is also related in this context, though the focus is on the costs of advertising to reveal quality by high-quality suppliers (e.g., see Kihlstrom and Riordan [1984] or Milgrom and Roberts [1986]).

are also related—they emphasize the benefits of firms in manipulating incentives because of joint production, though their story is noninformational and static.

Finally, also closely related to our article is the work by Hart et al. (1997), which uses the incomplete contracts approach to explain why governments run prisons and provide a definition of the “proper scope of governments.” With private ownership, managers receive a greater share of the gains they create, but this also induces them to engage in too much cost cutting at the expense of quality. We share with this article the emphasis on the potential costs of high-powered incentives associated with private ownership, but in our setup, these incentives arise not because of bargaining between the government and managers, but from the career concerns of producers, and different ownership structures affect incentives by influencing information transmission and the degree of career concerns.<sup>4</sup>

The rest of the article is organized as follows. Section 2 describes the environment and characterizes optimal incentives in a simple mechanism design problem. Sections 3–5 compare the incentive structure under markets, firms, and governments. Section 6 provides empirical evidence in support of the predictions of the model regarding government involvement in education and pension funds, while Section 7 concludes.

## 2. Model

We outline a simple two-period model to present our main intuition and focus on the teaching example for concreteness. An earlier version of this article (Acemoglu et al. 2003) presented the same analysis in the context of an infinite horizon model. Here, we focus on the two-period model to simplify the exposition.

### 2.1 The Environment

Consider a two-period economy with  $n$  teachers and  $2n'$  parents, where  $n' > n$ . We denote the set of teachers by  $\mathcal{T}$ . In each period,  $n'$  parents seek to obtain education for their children. A teacher  $i$  lives for both periods, and every period she can teach a single child.

The human capital of a child depends on the ability of the teacher assigned to him,  $a_t^i$ , and the unobservable “good” effort,  $g_t^i$ , put into teaching. We denote the human capital of a child assigned to teacher  $i \in \mathcal{T}$  by

$$h_t^i = a_t^i + f(g_t^i), \quad (1)$$

whereas a child not assigned to a teacher has human capital  $h_t = 0$ .

Ability  $a_t^i$  for teacher  $i \in \mathcal{T}$  is not known by any agent in the economy, but teachers and parents share the common belief that for each  $i \in \mathcal{T}$ , ability at

---

4. In this context, see also Acemoglu and Verdier (2000) and Prendergast (2003) on how government intervention or bureaucratic decision-making may create inefficiencies even when they are potentially improving the allocation of resources.

time  $t = 0$  is normally distributed with mean  $m_0^i$  and variance  $v_0$ , that is,  $a_0^i \sim \mathcal{N}(m_0^i, v_0)$ . Moreover, abilities are drawn independently across teachers. Teachers and parents also share the common knowledge that for each teacher  $i \in \mathcal{T}$ , ability at time  $t + 1$  is determined by  $a_{t+1}^i = a_t^i + \varepsilon_{t+1}^i$ , where  $\varepsilon_{t+1}^i$  is an i.i.d. teacher-level shock distributed as  $\mathcal{N}(0, \sigma_\varepsilon^2)$ . The function  $f(g)$  is increasing and strictly concave in  $g$ , with  $f(0) = 0$ . The level of  $h_t^i$  provided by a teacher is not observable to parents. Instead, parents rely on their child's test score  $s_t^i$ , which provides an imperfect signal of the true human capital accumulated by the child. The test scores are imperfect not only because they are noisy signals of human capital accumulation but also because they can be partially manipulated by a teacher. In particular, the test score  $s_t^i$  of a student assigned to a teacher  $i \in \mathcal{T}$  is given by

$$s_t^i = h_t^i + \gamma f(b_t^i) + \theta_t^i + \eta_t. \quad (2)$$

Here,  $b_t^i$  denotes the “bad” effort put in by the teacher and the parameter  $\gamma \geq 0$  reflects the influence of bad effort on test scores and thus captures the extent to which test scores can be manipulated by bad effort. In addition,  $\theta_t^i$  is an i.i.d. student-level shock distributed as  $\mathcal{N}(0, \sigma_\theta^2)$  (e.g., the ability of students to learn) and  $\eta_t$  is an i.i.d. common shock that *every* teacher receives in period  $t$ , distributed as  $\mathcal{N}(0, \sigma_\eta^2)$ . For example, if all students are given the same test,  $\eta_t$  can be thought of as the overall difficulty of the test and also captures any other cohort-specific difference in ability for performance.

The reason for calling the two types of efforts good and bad should be apparent now. Parents care about the good effort exerted by a teacher which influences the human capital of their child. However, they only observe the signal  $s$ , which can be manipulated by teacher's bad effort as well. In practice, bad effort may correspond to what is commonly referred to as “teaching to the test.” It involves rote learning, where a teacher just forces students to cram certain essential facts or methods, without explaining the concepts behind them or the connection between various facts and phenomena (see Hanaway 1992). Such cramming contributes less to human capital than good effort, but it serves to inflate their test scores. Bad effort might also be interpreted as teacher cheating, which improves test scores, but clearly has no beneficial effect on pupils' human capital (see the discussion in Section 6 below).

The expected utility of a parent at time  $t$  is given by

$$U_t^P = \mathbb{E}_t[h_t] - w_t,$$

where  $\mathbb{E}_t[\cdot]$  denotes expectations with respect to publicly available information at the beginning of time  $t$  and  $w$  is the wage paid to the teacher. This utility function implies that parents do not care about test scores directly but only about human capital.<sup>5</sup>

---

5. Our main results can be generalized to the case in which parents care about test scores as well as human capital.

Let  $t \in \{0, 1\}$  index the two periods. Expected utility of teacher  $i \in \mathcal{T}$  at beginning of period 0 is given by the time separable utility function

$$U_0^i = \mathbb{E}_0 \left[ \sum_{t=0}^1 \delta^t (w_t^i - g_t^i - b_t^i) \right],$$

where  $\delta < 1$  is her discount rate.

For now we focus on the case where  $n$  is very large, that is,  $n \rightarrow \infty$ , which allows us to ignore the common shock,  $\eta_t$  (since when  $n \rightarrow \infty$ ,  $\eta_t$  can be perfectly filtered out from average test scores, see below). The common shock will play an important role in Section 5 where we discuss incentives with government-provided education.

The exact timing of events is as follows:

1. In the beginning of period 0, each teacher is endowed with ability  $a_0^i \sim \mathcal{N}(m_0^i, v_0)$ .
2. Teachers potentially organize into firms/schools.
3. Parents compete by offering wages to teachers or firms, denoted by  $\{w_0^i\}_{i \in \mathcal{T}}$ .<sup>6</sup> These wages are not conditional on any future information.
4. At the end of period 0, human capitals,  $\{h_0^i\}_{i \in \mathcal{T}}$ , are produced for all pupils assigned to teachers and their test scores,  $\{s_0^i\}_{i \in \mathcal{T}}$ , are revealed.
5. Parents use the test score information to update their priors about the ability of teacher, obtaining updated beliefs  $a_1^i \sim \mathcal{N}(m_1^i, v_1^i)$  for each  $i \in \mathcal{T}$ .
6. Period 1 begins in the same sequence of events that take place.

We focus on the perfect Bayesian equilibrium of this game, where beliefs are derived from Bayes' rule given equilibrium strategies, and equilibrium strategies are optimal for teachers and parents at every information set given beliefs. We start by outlining how beliefs about teacher ability are updated.

## 2.2 Updating Beliefs

Recall that beliefs about teacher  $i \in \mathcal{T}$  at the beginning of first period can be summarized as  $a_0^i \sim \mathcal{N}(m_0^i, v_0)$ . Let  $S_0 \equiv [s_0^1 \cdots s_0^n]^T$  denote the vector of  $n$  test scores that parents observe during period 0 when each child is taught by a single teacher. Suppose that in equilibrium the effort levels of teacher  $i \in \mathcal{T}$  are  $g_0^i$  and  $b_0^i$ . Then, let  $Z_0 \equiv [z_0^1 \cdots z_0^n]^T$  such that

$$\begin{aligned} z_0^i &= s_0^i - f(g_0^i) - \gamma f(b_0^i), \\ &= a_0^i + \theta_0^i + \eta_0. \end{aligned}$$

Let  $a_1^i$  be the updated prior on teacher  $i$ 's ability conditional on observing  $Z_1$ . Then, Bayes' rule implies that  $a_1^i \sim \mathcal{N}(m_1^i, v_1)$ , where  $m_1^i$  and  $v_1$  denote mean and variance of posterior distribution. Moreover,

6. We are using  $\{w_0^i\}_{i \in \mathcal{T}}$  for wages for now with a slight abuse of notation, since when teachers form teams, parents will pay the price for their children to be taught by a team of teachers. See Section 4.

$$m_1^i = m_0^i + \beta(z_0^i - m_0^i) - \bar{\beta}(\bar{z}_0^{-i} - \bar{m}_0^{-i}), \quad (3)$$

where

$$\beta = \bar{\beta} = \frac{v_0}{(v_0 + \sigma_\theta^2)}, \quad (4)$$

$z_0^i$  is the  $i$ th element of the vector  $Z_0$  and refers to the signal from teacher  $i$ , whereas  $\bar{z}_0^{-i}$  is the average test score excluding teacher  $i$  and  $\bar{m}_0^{-i}$  is the average prior excluding teacher  $i$ . Since  $n \rightarrow \infty$ , we have  $(\bar{z}_0^{-i} - \bar{m}_0^{-i}) \rightarrow \eta_0$ , so the common shock is revealed and filtered out. The proof of equations (3) and (4) together with the expression for  $v_1$  is given in Appendix. It is also straightforward to see from Appendix that if  $a_0^i \sim \mathcal{N}(m_0^i, v_0^i)$ , our results generalize. Parents still filter out the noise from the common shock by using relative performance and the weight given to  $z_0^i$  while determining  $m_1^i$  becomes

$$\beta^i = \frac{v_0^i}{(v_0^i + \sigma_\theta^2)}$$

reflecting the heterogeneity in the precision about the beliefs regarding the ability of different teachers.

Equation (3) illustrates relative performance evaluation (yardstick competition) in the presence of the common shock  $\eta_0$ . The coefficient  $\bar{\beta}$  captures relative performance evaluation by emphasizing that an improvement in the score of a teacher creates a negative effect on the market's assessment of other teachers.

*Lemma 1.* Beliefs about the ability of teacher  $i \in \mathcal{T}$  are updated according to equations (3) and (4), where  $1 > \beta > 0$ . Moreover,  $\beta$  is increasing in  $v_0$  and decreasing in  $\sigma_\theta^2$ .

*Proof.* See Appendix.

The intuition behind the last part of Lemma 1 is that any increase in the variance of  $\theta$ ,  $\sigma_\theta^2$ , increases the noise in the signal, and makes it less valuable, and hence reduces  $\beta$ . An increase in  $v_0$  makes the signal more valuable. In other words, a greater  $\sigma_\theta^2$  relative to  $v_0$  implies that a given variation in test scores is less likely to come from teacher ability, so parents put less weight on differences in test scores in updating their posterior about teacher ability.

### 2.3 Efficient Allocations

Before characterizing the perfect Bayesian equilibria, we look at efficient allocations, which maximize “social welfare.” We define social welfare,  $U^W$ , as the sum of teachers’ and parents’ utilities. Since ability of teacher  $i$  enters additively in the utility function of parents, all teachers should choose the same effort level in a given period, which we denote by  $(g_t, b_t)$ . Let us focus on the case  $n \rightarrow \infty$ , so that there is no uncertainty about average realizations. Then, social welfare can be written as

$$U^W = \sum_{t=0}^1 \delta^t \bar{A} + f(g_t) - g_t - b_t, \quad (5)$$

where  $\bar{A}$  is the average ability of teachers in the population, and is a constant as  $n \rightarrow \infty$ .

*First-Best.* Maximizing equation (5) gives us the first-best. In the first-best, there is no bad effort,  $b_t = 0$ , and the level of good effort,  $g^{\text{FB}}$ , is given by  $f'(g^{\text{FB}}) = 1$  in both periods.

*Second-Best.* Since teacher effort and level of human capital are not directly observable, a more useful benchmark is given by solving for the optimal mechanism given these informational constraints. Solving the model backwards, we first see that there will never be any incentive for teachers to exert effort in the last period. This is due to the fact that effort is unobservable and wages are not conditioned on future realizations of pupils' human capital (recall the timing of events at the end of Section 2.1). Consequently,  $g_1^{\text{SB}} = b_1^{\text{SB}} = 0$ , and the human capital output of teacher is  $h_1^i = a_1^i$ . Given the timing of events we assume, this feature holds with all organizational forms, and throughout, we therefore focus on the first period where teacher's effort can be manipulated by changing her incentive scheme,  $w_1^i$ , in the second period.

Recall that wage  $w_1^i$  can be a function of the available information at the beginning of the second period. The information set relevant for teacher  $i \in \mathcal{T}$  can be defined as  $\Omega_1^i = [m_1^i, s_0^i, s_0^{-i}]$ . Note that  $\Omega_1^i$  contains the test score associated with teacher  $i$  and the vector of test scores for all other teachers,  $s_0^{-i}$ , as well as the updated beliefs regarding the ability of teacher  $i$ ,  $m_1^i$ .<sup>7</sup> Therefore, for characterizing the second-best, there is no loss of generality in focusing on  $\Omega_1^i$ . Let  $\omega_1^i(\Omega_1^i)$  be the wage paid to teacher  $i$  in period 1. Then, the constrained maximization problem to determine the second-best allocation of effort in period 0 can be written as:

$$\begin{aligned} & \max_{\{w_1^i(\Omega_1^i)\}} (\bar{A} + f(g_0) - g_0 - b_0) \\ & \text{where } \{g_0, b_0\} \in \arg \max_{\{g_0', b_0'\}} (\delta \mathbb{E}_0[w_1(\Omega_1^i)] - g_0' - b_0'). \end{aligned} \quad (6)$$

We leave details of the maximization problem (6) to Appendix. An important consequence of the incentive compatibility condition in equation (6) is that any effort combinations  $\{g_0, b_0\}$  must satisfy

$$\gamma f'(b_0) = f'(g_0).$$

This equation shows that teachers can be encouraged to exert good effort only at the cost of bad effort. As a result, the opportunity cost of inducing high effort is greater in the second-best problem than in the first-best.

7. It is straightforward to see that  $\Omega_1^i$  is the largest set of contractible information about teacher  $i$ ; if teacher  $i$  is part of a nonsingleton team in first period, then there will be less information about his ability.



Next consider a wage schedule of the form  $w_1^i = \alpha m_1^i + \kappa$ , which links teacher compensation to their contemporaneous perceived ability (see Appendix to see why focusing on such linear contracts is without loss of generality). This formulation of incentives is similar to the seminal career concerns article by Holmstrom (1999). The extra effort put in by a teacher in period 0 increases her test score in period 0. There are no immediate rewards for this increase as the teacher has already been paid her wage. However, an increase in test score raises her perceived ability in period 1 due to the updating rule (3). This marginal benefit of higher test scores in period 0 can be summarized as  $\delta\alpha\beta$ : greater  $\beta$  implies that teacher effort will have a larger influence on future perceptions of her ability, and thus greater future rewards (which is the reason why discount factor  $\delta$  also matters). We define  $\beta$  as the “career concerns coefficient.” The marginal benefit is also increasing in  $\alpha$ , which can be thought of as “the market-reward coefficient”—how much the market rewards a unit increase in the perceived ability of the teacher.

The marginal benefit of higher test scores implies that privately optimal levels of good and bad effort are

$$f'(g_0) = \gamma f'(b_0) = \frac{1}{\delta\alpha\beta}.$$

This implies that a greater  $\alpha$ , that is, higher-powered incentives, translate into greater good and bad effort, and for the reasons explained in the previous paragraph, the magnitude of this effect depends on both the career concerns coefficient  $\beta$  and the discount factor  $\delta$ .

The following proposition, which is proved in Appendix, characterizes the second-best effort levels and determines the value of  $\alpha$  that will induce these effort levels.

*Proposition 2.* The second-best solution is given by  $g_0 = g^{\text{SB}}$ , and  $b_0 = b^{\text{SB}}$ , with  $g^{\text{SB}} < g^{\text{FB}}$ . When each child is taught by a single teacher, the optimal wage schedule is given by  $w_1^i = \alpha^{\text{SB}} m_1^i + \kappa$ , where

$$\alpha^{\text{SB}} \equiv \frac{1}{\delta\beta f'(g^{\text{SB}})}, \quad (7)$$

and for any nonnegative  $\kappa$ . Both  $g^{\text{SB}}$  and  $\alpha^{\text{SB}}$  are monotonically decreasing in  $\gamma$ . Moreover, there exists a threshold  $\underline{\gamma}$ , such that  $\alpha^{\text{SB}} < 1$ , for  $\gamma > \underline{\gamma}$ .

Proposition 2 highlights the trade-off that a social planner faces given the informational constraints. The planner needs to provide incentives to teachers in order to induce effort. However, high-powered incentives lead to both good and bad effort. This association between good and bad effort increases the shadow cost of increasing good effort, leading to a lower level of good effort in the second-best relative to the first-best. The parameter  $\gamma$  captures the cost of higher incentives in the form of bad effort. Hence, an increase in  $\gamma$  increases

the scope for bad effort and reduces the second-best level of good effort,  $g^{\text{SB}}$ , and consequently, the optimal level of incentives for the teacher,  $\alpha^{\text{SB}}$ .

Expression (7) shows that for a given  $\beta$ , the market-reward coefficient,  $\alpha$ , can be adjusted to give us  $g^{\text{SB}}$ . Similarly, if it were possible to manipulate  $\beta$  (as different organizational forms will do below), the second-best effort can also be achieved by changing the value of  $\beta$  for a fixed level of  $\alpha$ . This is stated in the next corollary (proof omitted).

*Corollary to Proposition 2.* The second-best equilibrium can alternatively be described by fixing  $\alpha$  and setting the career concerns coefficient on an individual teacher's test score equal to

$$\beta_{\text{SB}} \equiv \frac{1}{\delta \alpha f'(g^{\text{SB}})}. \quad (8)$$

This preceding discussion therefore highlights two different channels via which the second-best allocation can be obtained. The first is by manipulating  $\alpha$ , that is, how the market rewards "success," and the second is by manipulating  $\beta$ , that is, the teachers' career concerns. In the sections that follow, we discuss how successful different organizational forms are in manipulating the career concerns coefficient to improve the allocation of resources.

### 3. Incentives in Markets

In this and the next two sections, we consider three different organizational structures—markets, firms, and governments—and compare the incentives they provide to teachers. Consider first the simplest model of perfectly competitive markets. Every teacher works independently, teaches a single child, and sells her teaching services in the market every period. This economy therefore corresponds to the timing of events given in Section 2.1. Bertrand competition among parents implies that wage  $w_1^i$  for teacher  $i \in \mathcal{T}$  is given by

$$w_1^i = m^i. \quad (9)$$

The market equilibrium is therefore similar to the second-best equilibrium, except that now  $\alpha$  is fixed to be 1. This leads to the following result.

*Proposition 3.* The market equilibrium is characterized by good effort level  $g^M$  in the first period, where

$$f'(g^M) = \frac{1}{\delta \beta}.$$

We have that  $g^M < g^{\text{SB}}$  if  $\gamma < \underline{\gamma}$ , and  $g^M > g^{\text{SB}}$  if  $\gamma > \underline{\gamma}$ .

The proof follows immediately from Proposition 3 and is omitted. The result that  $g^M < g^{\text{SB}}$  if  $\gamma < \underline{\gamma}$  is similar to the result in Holmstrom (1999) that, with discounting, career concerns are typically insufficient to induce the optimal level of effort. So in this case, even markets do not provide strong enough incentives. There may be certain nonmarket institutions (e.g., tournaments) that strengthen

incentives even further, though we do not focus on those here.<sup>8</sup> Therefore, when  $\gamma < \underline{\gamma}$ , markets are the preferred form of organization. This leads to the conclusion, mentioned above, that where quality concerns are unimportant relative to the total amount of effort/investments, services should be sold in markets.

The case where  $\gamma > \underline{\gamma}$ , on the other hand, leads to the opposite conclusion. Now, the natural career concerns provided by the market equilibrium create too high-powered incentives relative to the second-best. The extent to which the market provides excessively high-powered incentives depends on the career concerns coefficient,  $\beta$ , and via this, on  $\sigma_0^2$  and  $v_0$ . When  $\sigma_0^2$  is small relative to  $v_0$ ,  $\beta$  is high, and teachers in the market care a lot about their pupils' scores, giving them very high-powered incentives. In this case, since markets are encouraging too much bad effort, firms or governments may be useful by modifying the organization of production to dull incentives. We next turn to a discussion of the role of firms and governments in providing appropriate incentives when markets lead to too high-powered incentives, that is, when  $\gamma > \underline{\gamma}$ .

#### 4. Incentives in Firms

So far we have assumed that each teacher works independently. We now allow teachers to be organized into firms, whereby teacher  $i \in \mathcal{T}$  works for firm or "school"  $j$ . Since we focus on the case where  $\gamma > \underline{\gamma}$ , market incentives are too high powered. Firms might be able to dampen the power of incentives by creating teams of teachers and weakening the signaling ability of individual teachers, which could potentially improve the allocation of resources.<sup>9</sup>

We model firm  $j$  as a partnership of  $K^j$  teachers working together to jointly teach  $K^j$  children. The human capital,  $h_t^j$ , of a child studying in firm  $j$  is therefore given by

$$h_t^j = \bar{a}_t^j + \overline{f(g_t)}^j, \quad (10)$$

where  $\bar{a}_t^j = \frac{1}{K^j} \sum_{i \in \mathbb{K}^j} a_t^{ij}$  and  $\overline{f(g_t)}^j = \frac{1}{K^j} \sum_{i \in \mathbb{K}^j} f(g_t^{ij})$ , where  $\mathbb{K}^j$  denotes the set of teachers in firm  $j$ . Given the joint production of teaching inside a firm, test score of a child in firm  $j$  is given by

$$s_t^j = h_t^j + \gamma \overline{f(b_t)}^j + \bar{\theta}_t^j + \eta_t, \quad (11)$$

where  $\overline{f(b_t)}^j$  and  $\bar{\theta}_t^j$  are defined analogous to  $\bar{a}_t^j$  as averages over the set of teachers in  $\mathbb{K}^j$ . A critical implication of equation (11) is that parents only

8. One can also imagine organizations that reward teachers according to a wage function along the lines of  $w_t^i = \alpha m_t^i + \kappa$  with  $\alpha > 1$  to strengthen incentives beyond those provided by the market. Firms, modeled below as teams of teachers, are unable to do so, however, since the "balanced budget" requirement imposes that  $\alpha \leq 1$  and  $\kappa \geq 0$ . See Holmstrom (1979).

9. We limit the analysis to the implicit incentives provided by firms. In addition, firms could improve the allocation of resources by providing explicit incentives, that is, writing contracts with their employees that are perfectly observed by their customers. Although this is a possibility in the symmetric information case, in the asymmetric information case, which is our main focus, such contracts are not useful, since the firm can write additional side contracts, not observed by the customers, changing their employees' incentives away from those implied by the explicit contracts.

observe the aggregate or average test score of all the teachers (or pupils) in the firm. Therefore, firms have the ability to shut down individual signals (test scores) of teachers.

We first take the set of firms as given and characterize the equilibrium. In this case, the timing of events is very similar to that given at the end of Section 2.1, except that only average test scores for firms are publicly observed. Subsequently, we allow firms to form endogenously.

There are a total of  $J$  firms in the economy with  $\sum_{j=1}^J K^j = n$ , each with  $K^j < \infty$ . As  $n \rightarrow \infty$ , we also have  $J \rightarrow \infty$ . Consequently, along the equilibrium path parents can again back out the signal  $\bar{z}_0^j = \bar{a}_0^j + \bar{\theta}_0^j + \eta_0$  from  $\bar{s}_0^j$ . Let

$$\bar{m}_t^j = \frac{1}{K^j} \sum_{i \in \mathbb{K}^j} m_t^{ji} \quad (12)$$

be the expected ability of the teachers in firm  $j$  at time  $t$ . Then, parents update their time  $t = 1$  belief about teacher  $i$ 's (working in firm  $j$ ) ability according to an updating formula similar to equation (3):

$$m_1^{ji} = m_0^{ji} + \beta_F(\bar{z}_0^j - \bar{m}_0^j) - \bar{\beta}_F(\bar{z}_0^{-j} - \bar{m}_0^{-j}). \quad (13)$$

Although parents can only observe the average test score of all the teachers in the firm, it is possible for those inside the firm to have more information about each individual teacher's performance. Thus, in addition to the average test score in the firm, teachers employed in the firm and the owner also observe the following signal of each teacher's performance (test score):

$$s_0^{ji} = a_0^{ji} + f(g_0^{ji}) + \gamma f(b_0^{ji}) + \theta_0^{ji} + \tilde{\theta}_0^{ji} + \eta_0, \quad (14)$$

where  $\tilde{\theta}_0^{ji}$  is a normal error term, distributed as  $\mathcal{N}(0, \sigma_\theta^2)$ . When  $\sigma_\theta^2 \rightarrow \infty$ ,  $\tilde{\theta}_0^{ji}$  has a very large variance and insiders have exactly the same information as outsiders—that is, there is no asymmetric information. We start with this case of no asymmetric information and later analyze the case where insiders have better information.

Bertrand competition between parents implies that a group of teachers are paid their expected contribution to human capital. Thus, the *average* earnings of a teacher in firm  $j$  at time  $t \in \{0, 1\}$  is

$$\bar{w}_t^j = \bar{m}_t^j + \overline{f(g_t)}^j, \quad (15)$$

where  $\bar{m}_t^j$  is given by equation (12), and total revenue of firm  $j$  is  $K^j \bar{w}_t^j$ .

We also need to know how each individual teacher is rewarded (i.e., how the total revenue  $K^j \bar{w}_t^j$  is divided between the teachers). We assume that each teacher's wage at time  $t \in \{0, 1\}$  is given by

$$w_t^{ji} = m_t^{ji} + f(g_t^{ji}), \quad (16)$$

where  $m_t^{ji}$  is the expected ability of teacher  $i$  in firm  $j$  given the insiders' information set, with evolution given by equation (13). This wage rule parallels the

market wage rule, (9), thus making it clear that the advantage of firms does not come from manipulating the wage rule but from obscuring information.<sup>10</sup> In fact, the firm as an entity has exactly the same high-powered incentives as the individual teachers in the previous section; this observation will play an important role in limiting the ability of firms to manipulate the power of incentives.

We also assume that the set of teachers in firm  $j$ ,  $\mathbb{K}^j$ , is chosen at time 0 and is not changed thereafter. In other words, teachers do not switch “teams” after initial assignment.<sup>11</sup> Finally, we assume that the firm (partnership) maximizes net revenue per partner (teacher), where net revenue represents revenues minus the effort costs of a teacher. This implies that the partnership takes the optimal behavior of each teacher derived from the equilibrium wage schedule (16) as given, and maximizes expected time  $t = 0$  average net benefits to its members. As before, no teacher will invest any effort in the last period. Moreover, all teachers within the team will choose the same level of good and bad effort, denoted by  $g_0^j(K^j)$  and  $b_0^j(K^j)$  for a team of size  $K^j$ . Consequently, the maximization problem for firm  $j$  that determines its size can be written as

$$\max_{K^j} \mathbb{E}_0^j[(\bar{w}_0^j(K^j) + \delta \bar{w}_1^j(K^j) - g_0^j(K^j) - b_0^j(K^j))] \quad (17)$$

subject to

$$\{g_0^j(K^j), b_0^j(K^j)\} \in \arg \max_{\{g_0^j, b_0^j\}} \delta \mathbb{E}_0[w_1^{jj}(K^j)] - g_0^{jj} - b_0^{jj},$$

where  $\bar{w}_t^j$  is given by equation (15),  $K^j$  denotes the total number of teachers in that firm, and  $\mathbb{E}_0^j$  is the expectations given the information set of the insiders in

10. Similar results would apply if wages within the partnership were determined by some type of bargaining instead of the wage equation (16).

11. It can be shown that, as long as we are in the case with  $\gamma > \underline{\gamma}$ , “no-switching” is an equilibrium. To see this briefly, consider a symmetric equilibrium. According to equation (16), every teacher is paid her expected output equal to  $w_t^{jj} = m_t^{jj} + f(g_t^{jj})$ . The expected utility of a teacher if she stays in the firm is therefore given by  $U^{jj} = \mathbb{E}_0[(m_0^{jj} + f(g_0) - g_0 - b_0 + \delta m_1^{jj})]$ . Since  $\mathbb{E}_0^j[m_1^{jj}] = m_0^{jj}$ , we have  $U^{jj} = (1 + \delta)m_0^{jj} + f(g_0) - g_0 - b_0$ . We next discuss deviations to switch to another team and to opening a singleton firm (entrepreneurship), starting with the latter. Compute the switcher’s utility assuming that in all future periods, he/she is expected to, and will, exert good and bad effort equal to  $g$  and  $b$ . In this case, we continue to have  $\mathbb{E}_0[m_1^{jj}] = m_0^{jj}$ , and as long as singleton firms are not of the optimal size, there will be a loss of utility for the switcher. In addition, after switching, market perceptions of his/her ability will be negatively correlated with those of his/her old coworkers. This will induce the teacher who switches to put in more effort (both good and bad). As long as  $\gamma > \underline{\gamma}$ , this will be rewarded by the market less than the cost of effort, and hence greater effort will reduce the utility of the switching teacher.

Next, we verify that when no other teacher switches teams, a deviation to switching to another team by a single teacher is also not profitable. To see this, note that the payment to a switcher in the new firm will be according to the public perception of his/her ability,  $m_1^{jj}$ , which can be more than  $\mathbb{E}_1^j[m_1^{jj}]$ , the expectation of ability given the firm  $j$  information set, introducing an adverse selection problem. Although a full analysis of adverse selection in this context is beyond the scope of our article, we can see that since there is no switching, a reasonable set of off-the-equilibrium path beliefs would be that switchers are selected from those with the most to gain from a deviation. This would imply that switchers have arbitrarily low  $\mathbb{E}_1^j(m_1^{jj})$ , making no-switching an equilibrium.

firm  $j$  at time  $t = 0$ .  $\bar{w}^j$ ,  $g$ , and  $b$  are written as functions of  $K^j$  to emphasize that the size of the firm will influence incentives and payments.

#### 4.1 Symmetric Information— $\sigma_0^2 \rightarrow \infty$

In this case, the firm can only make payments to teachers conditional on the average signal from all the teachers  $\bar{s}_t^j$ . To see the benefits of large firms in the simplest possible way, let us return to the updating equation, (13), which is similar to the updating equation in the market case, (3). The career concerns coefficient for an individual teacher is different, however. In particular, in a firm of size  $K$ , the individual career concerns coefficient is  $\beta_F/K$ . The reason for this decline is the “moral-hazard-in-teams” problem. For each incremental increase in her test score, a teacher only gets rewarded for a fraction  $1/K$  of the value created for the team. Moreover, as the proof to Proposition 3 in Appendix shows,  $\beta_F = \beta$  and  $\bar{\beta}_F = \bar{\beta}$ . Since  $\beta/K$  is decreasing in  $K$ , the power of incentives can be reduced by increasing firm size, and in the case where  $\gamma > \underline{\gamma}$ , there exists a  $K^*$  such that  $K^* \equiv \beta/\beta_{SB}$ , where  $\beta_{SB}$  is the career concerns coefficient that would ensure the second-best with  $\alpha = 1$ , as defined by equation (8).

We can now see that the solution to the second-best problem (6) can be implemented as a competitive equilibrium among firms that maximize equation (17). As the wage function in equation (9) highlights, competition among firms implies that  $\alpha = 1$ . It then follows from Corollary to Proposition 2 that firm  $j$  will expand size  $K^j$  until the power of incentives is given by  $\beta_{SB}$ , that is, until  $\beta/K^j = \beta_{SB}$ . This of course implies that each firm will select  $K^j = K^*$ .<sup>12</sup>

*Proposition 4.* Suppose that  $\sigma_0^2 \rightarrow \infty$ . Then, for a firm of size  $K^j$ , the good effort level chosen by a teacher,  $g_0$ , is given by  $g_0^F(K^j)$ , where  $g_0^F$  is monotonically decreasing in  $K^j$  with  $g_0^F(1) = g_0^M$  and  $g_0^F(K^j) \rightarrow 0$  as  $K^j \rightarrow \infty$ .

When  $\gamma > \underline{\gamma}$ , there exists a unique equilibrium where firms have size equal to  $K^* \equiv \beta/\beta_{SB} > 1$  and where teachers exert the second-best level of good effort,  $g_0^{SB}$  in the first period.

As in the market equilibrium, a teacher is still paid her expected output. However, the marginal effect of test score at time 0 on future expected ability is lower in firms than in markets. In other words, firms lower the career concerns coefficient from  $\beta$  to  $\beta/K$ , thus weakening individual incentives.

The reduction of career concerns effects under firms thus redresses the “over-incentivization” problem. As firms compete to maximize their value, they endogenously expand to a size of  $K^*$  and second-best allocation of Section 2.3 is achieved.

#### 4.2 Asymmetric Information and No Commitment— $\sigma_0^2 < \infty$

The preceding analysis highlighted how organizational structure of firms can be used to suppress information. The question still remains, as to what extent

12. Here, we ignore “integer issues” since we are focusing on the case where Otherwise,  $K^j = K^*$  would not be possible for all  $j$ , and at least one firm may have to have a different size.

a firm as a whole (or the principal/owner) has access to information regarding an individual teacher's test score. Proposition 4 above assumed that no one inside the firm can observe a teachers' test score either.

We therefore relax the assumption of symmetric information and assume that  $\sigma_\theta^2 < \infty$ . This implies that insiders (the "partnership") now observe a noisy signal of individual teacher performance as well as the public signal coming from average firm performance. In addition, we follow the collusion literature and assume that firms can enter into side deals with their employees. This implies that if they announce some wage schedule for their employees, they do not necessarily have to stick to these.<sup>13</sup> This is plausible given the various ways in which firms can reward their employees, without outsiders detecting the exact form of contractual arrangements. Without the asymmetry of information, firms had no ability to manipulate employee rewards, which would be given by equations (13) and (16). Thus, there was no need for firms to commit to wage contracts. The commitment problem—the inability of firms to refrain from giving high-powered incentives to their employees—arises due to asymmetric information.

The commitment problem implies that firm owners (or firms) would like to use their private information to provide higher incentives through side contracts with their teachers. This result is summarized in the next proposition.

*Proposition 5.* Suppose that  $\gamma > \underline{\gamma}$  and  $\sigma_\theta^2 < \infty$ . Then there exists  $\overline{\sigma_\theta^2}$ , such that when  $\sigma_\theta^2 > \overline{\sigma_\theta^2}$ , there is a unique equilibrium in which firms have size equal to  $K^{**}(\sigma_\theta^2) > 1$ , where  $K^{**}(\sigma_\theta^2)$  induces the second-best level of effort  $g^{SB}$ , and is decreasing in  $\sigma_\theta^2$ . When  $\sigma_\theta^2 \leq \overline{\sigma_\theta^2}$  the second-best outcome cannot be achieved. When  $\sigma_\theta^2 = 0$ , the firm equilibrium leads to the market outcome, that is, the good effort level  $g^M$ .

*Proof.* See Appendix.

Let us illustrate the intuition for this result using the case of complete asymmetric information that is  $\sigma_\theta^2 = 0$ . Suppose it were possible for firm  $j$  to choose a size  $K^*$  that implements the second-best level of effort. Recall from Section 3 that the outside market uses a career concerns coefficient of  $\beta_F = \beta$ , with  $\beta$  defined by equation (4). This implies that the marginal benefit for the firm (partnership) to increase average test scores is exactly the same as the marginal benefit of increasing individual teacher's test score in the market setup. Moreover, since  $\sigma_\theta^2 = 0$ , insiders observe teachers' exact test score and can thus promise to reward them according to a career concerns coefficient of  $\beta$  as in the market

13. As in the collusion literature (e.g., Tirole [1986]), this assumption raises the question of how firms (or partnerships) can credibly form such side deals. Like the wage equation (16), this result can be derived as a repeated-game equilibrium in an infinite horizon setting, like the one considered in the previous version of the article, Acemoglu et al. (2003). Models of side deals and collusion in infinitely repeated relationships are further discussed in Acemoglu (1996) and Martimort (1999).

organization, which would increase the average net revenue of the partnership. The firm will therefore give full market incentives to the teachers inside, contradicting our assumption that  $K^*$  implemented second-best effort.

Thus, choosing a firm size of  $K^*$  (as given by Proposition 3) is no longer a credible commitment to low-powered incentives and to the second-best level of good effort. Instead, the strength of incentives will be determined by the amount of information the firm has about each employee's performance. Since the firm's (the insiders') information about individual performance is also imperfect, that is, typically,  $\sigma_\theta^2 > 0$ , average performance of the firm is still informative about each employee's ability. Therefore, firm size, by affecting how informative average performance is about individual ability, still influences how powerful each employee's incentives are. Generally, the larger the size of the firm, the less information there is about an individual's performance inside the firm and the less powerful are equilibrium incentives. Therefore, a firm might still be able to credibly commit to low-powered incentives by further increasing its size to  $K^{**}(\sigma_\theta^2)$ , thus reducing teachers' incentives even after taking into account the ex post manipulation of these incentives. Nevertheless, the precision of internal signals puts a lower bound on how much the firm can dull incentives through "team production." In particular, if  $\sigma_\theta^2 \leq \sigma_\theta^2$  for some critical threshold  $\sigma_\theta^2$ , then there is sufficiently good internal information about teacher performance that even a very large firm would not be able to dull incentives sufficiently.

The intuition for why asymmetric information and associated commitment problems make firms less useful can be alternatively described as follows: when production is organized within firms, individual teachers have relatively weak incentives because of the moral-hazard-in-teams problem. The firm as an entity, or its owner, however, has strong incentives, since it is the residual claimant of profits. These incentives *trickle down* to employees if a firm has better information on individual teacher's performance than outside world.

## 5. Incentives in Governments

Let us now imagine a world with  $\gamma > \underline{\gamma}$  and  $\sigma_\theta^2$  small so that markets provide too high-powered incentives, and firms cannot commit to dulling individual incentives because of asymmetric information. For simplicity, let  $\sigma_\theta^2 = 0$ , so firm and market equilibria are identical with teacher exerting too much effort. The question we investigate is whether governments can help improve allocation in this case.

Governments are widely believed to offer "flat" incentives. Empirical studies suggest considerable wage compression in governments relative to the private sector (e.g., Johnson and Libecap 1994). Civil service rules in many countries make firing difficult and tightly link pay with education and seniority. The literature discusses a variety of reasons for low-powered incentives in governments ranging from the absence of market discipline (Niskanen 1971; Hanushek 1996) to an optimal design to avoid collusion and corruption (e.g., Crozier 1967; Tirole 1986; Banerjee 1997; Acemoglu and Verdier 2000). In our model, we can think of these concerns imposing a wage structure on government organizations of the



form  $w_t^j = \alpha^G m_1^j + \kappa$ . If  $\alpha^G$  were close to  $\alpha^{\text{SB}}$ , that is, if incentives in government-run firms were close to the power of incentives necessary to achieve the second-best, government organization would be useful.

Aside from this possibility, our model suggests other reasons why governments may be particularly able to commit to low-powered incentives. We now discuss these issues using a stylized model of government organization. This setup is similar to the case of firms, with teacher  $i$  teaching in firm  $j$ . However, the critical difference is that all firms are now controlled and operated by a government politician. The politician decides the size of schools as well as individual teacher rewards. The politician is also self-interested and similar to the objectives of individual teachers; she would like to convince the market (in particular, voters) that she has high ability in order to increase her reelection probability.

We denote the politician's true ability and voters' perception of this at time  $t$  by  $a_t^p$  and  $m_t^p$ , respectively, where initial politician's ability  $a_0^p$  is distributed as  $\mathcal{N}(m_0^p, v_0^p)$ . In particular, we assume that the politician has an objective function given by

$$U_0^{\text{pol}} = \mathbb{E}_0 \left[ \sum_{t=0}^1 \delta^t (m_t^p - C_t) \right], \quad (18)$$

where  $C_t$  is the cost per student of the schooling system, and is given by  $C_t = \bar{w}_t$ , where  $\bar{w}_t$  is the average teacher wage. This utility function implies that the politician always likes to convince parents (or the voters) that he has high ability and faces a cost in terms of the expenditures on the education budget in this process.

Given this setup, we discuss reasons why governments may be better able to commit to low-powered incentives than markets and firms. First, even when the politician has an incentive to inflate test scores, these incentives may be lower than those in firms in the presence of common shocks because of the absence of yardstick competition in governments. Second, even in the absence of common shocks, governments may still have lower career concerns because of the reelection uncertainty that politicians face driven by factors unrelated to their performance in this sector.

### 5.1 Government Operation with Common Shocks

In order to make the case of governments interesting, we need to allow room for the actions of the politician to influence outcomes other than through her effect on teacher effort. In general, decisions taken by education ministers or prime ministers can have important influences on aggregate outcomes, for example, through teacher selection, by affecting incentives in other dimensions, or by influencing the curriculum. We allow for this possibility in a simple way by assuming that the ability of the politician also matters for the human capital attained by the children. In particular, assume that the human capital of a student taught by teacher  $i$  is

$$h_t^i = a_t^i + \lambda a_t^p + f(g_t), \quad (19)$$

where  $a_t^p$  is the ability of the politician in charge of the schooling system. Politician's ability now influences the human capital of all the children in the school system because of some other dimension of incentives that the politician provides to teachers, or because of her decisions. Consequently, the politician has an incentive to inflate test scores in order to improve others' perception of her ability. The preceding analysis is unaffected by this modification and the term  $\lambda a_t^p$  is now included in an augmented common shock,  $\eta_t' = \lambda a_t^p + \eta_t$ .

What we want to point out, however, is that even in this case the government may have a comparative advantage in providing low-powered incentives. When an individual school inflates its own test scores, this has a negative effect on other schools because of the relative performance evaluation used by the market to remove the effect of the common shock,  $\eta_t$ . This intensifies the negative externality and encourages private schools to give high-powered incentives to their teachers. In contrast, with government operation, the politician is in charge of the whole school system, so when citizens (voters) update their beliefs about the ability of the politician, the common shock is not filtered out and acts as an additional source of noise, thus *weakening* the incentives of the politician.

More formally, parents (or voters) observe all test scores at the end of first period and update their beliefs regarding the ability of the politician according to the equation

$$m_1^p = m_0^p + \beta^p(z_0 - \lambda m_0^p), \quad (20)$$

where

$$z_0 = \frac{1}{J} \sum_{j=1}^J \bar{s}_0^j - \bar{A} - \overline{f(g_0)} - \overline{\gamma f(b_0)} = \lambda a_0^p + \eta_0, \quad (21)$$

and

$$\beta^p = \frac{\lambda v_0^p}{\lambda^2 v_0^p + \sigma_\eta^2}, \quad (22)$$

where  $J$  is the number of firms in the economy,  $\bar{A}$  is the average ability of teachers in the population, and  $\bar{s}_t^j$  refers to the average test score of firm  $j$  at time  $t$ . These updating equations have an intuition similar to equations (3) and (4). The updating is now about the ability of the politician. For updating, only the average test score in the population is relevant, and in equilibrium, this average test score is equal to  $\bar{A} + \overline{f(g_t)} + \overline{\gamma f(b_t)} + \lambda a_t^p + \eta_t$ . The career concerns coefficient of the politician,  $\beta^p$ , is different from that of firms (or individual teachers),  $\beta$ , because learning now is about the ability of the politician, which may have a different distribution, and more importantly, because noise comes from the aggregate shock,  $\eta_t$ , not from the student performance shocks, the  $\theta_t$ 's. The reason why  $\sigma_\eta^2$  did not feature in the updating

equations (3) and (4) is that relative performance evaluation eliminated this aggregate shock. With government operation, relative performance evaluation is not possible, since everything is run by the government,<sup>14</sup> and this makes (the perception of) government performance dependent on the realization of the aggregate shock. As a result, the politician receives credit for only part of the improvements in test scores, weakening her incentives, and therefore, indirectly those of the whole government organization. The greater the  $\sigma_\eta^2$ , that is, the more important the aggregate shock, the smaller the  $\beta^p$ , and the weaker the incentives in governments. In the limit, as  $\sigma_\eta^2 \rightarrow \infty$ , the politician has completely flat incentives.

Next, we look at the equilibrium level of effort chosen by the teachers under government operation,  $g^G$ , which will be determined by the incentives trickling down to the individual teacher level. Given the politician's own incentives in equation (22), we can determine the wage schedule that the politician will offer to each of his teachers. In particular, assume that the politician offers each teacher a linear wage function of the form

$$w_1^i = \alpha_1^p m_1^i + \kappa_1,$$

where  $m_1^i$  is the expected ability of teacher  $i$  at time  $t = 1$  and  $\kappa_1$  is a constant. First, consider the case where the level of incentives provided to teachers  $\alpha_1^p$  is observable. Then, even though the politician can manipulate teacher incentives, she will receive no benefit from this, since voters will effectively observe the level of good and bad effort exerted by teachers. In this case, the results would be identical to those with no politician effects, and the politician would simply choose  $\alpha_1^p = \alpha^{SB}$  and achieve the second-best.

However, parallel to our treatment of firms where teacher incentives inside the firm are not observed by outsiders, it may be more reasonable to presume that  $\alpha_1^p$ 's are not observable citizens. Interestingly, even in this case, government operation provides weaker incentives than markets and firms. We now analyze this case by considering the maximization problem of the politician, which is to maximize equation (18) by choosing the wage function of teachers. Since the government acts as a monopolist, it will only give each teacher his/her minimum reservation utility. Let  $\underline{u}$  be the spot reservation utility of a teacher. Then, in order to induce efforts  $g_0$  and  $b_0$  in the first period, the government must pay each teacher a wage equal to  $w_0^i = \underline{u} + g_0 + b_0$  in the first period, and promise a wage  $w_1^i = \alpha_1^p m_1^i + \kappa_1$  in the second period such that incentive  $\alpha_1^p$  is high enough to induce the desired level of effort. The equilibrium level of  $\alpha_1^p$  (and hence effort level  $g^G$  in first period) will be determined by

---

14. This argument needs to be qualified when local politicians run local school districts, for example, as in the United States. In this case, there will be some amount of competition even with government operation. Nevertheless, given the importance of district-specific shocks, the extent of yardstick competition might be much less than the case of private ownership, with competition between private schools, thus qualitatively leading to the same type of comparison as that emphasized in this section.

the politician's objective function (18). Maximizing this objective function with respect to  $\alpha_1^p$  gives us the following result.

*Proposition 6.* Suppose that  $\sigma_\theta^2 = 0$  and  $\gamma > \underline{\gamma}$ , so that both markets and firms lead to the same inefficiently high level of effort  $g^M > g^{SB}$ , with  $g^M$  given in Proposition 3. The equilibrium level of effort under government operation,  $g^G$ , is then given by

$$f'(g^G) = \frac{1}{\delta\beta^p}.$$

We have  $g^G < g^M$  if and only if  $\beta^p < \beta$ . Moreover,  $\beta^p$  is decreasing in  $\sigma_\eta^2$  and increasing in  $\lambda$ .

*Proof.* See Appendix.

The proposition establishes that, because of the presence of common shocks, government operation often provides weaker incentives than firms and markets, even though politicians have an interest in inflating test scores, and the manipulation of teacher incentives by the politician is not observed by voters. The presence of common shocks increases the amount of noise in the performance of the politician, weakening her incentives. These weaker incentives then trickle down to the teachers.

More specifically, when  $\beta^p < \beta$ , government organization provides less high-powered incentives than markets and firms because the politician has less to gain by inflating test scores. This is likely to be the case when aggregate shocks are large, that is, when  $\sigma_\eta^2$  is large, and when the contribution of the politician to aggregate test scores,  $\lambda$ , and the room for the politician to prove that she has high ability,  $\sigma_p^2$ , are limited. This reasoning also suggests that government operation may be beneficial in reducing incentives in activities where there is more scope for unproductive signaling effort and politicians have limited room to manipulate aggregate performance to improve their standing. In contrast, when  $\sigma_\eta^2$  is small, and/or when  $\lambda$  and  $\sigma_p^2$  are large, politicians can manipulate incentives more than profit-maximizing firms, and government operation is likely to lead to a deterioration in the allocation of resources.

## 5.2 Government Operation Under Reelection Uncertainty

The above analysis may be criticized on the grounds that it is not government operation per se but monopoly that is essential to limit yardstick competition in the presence of common shocks.<sup>15</sup> If so, perhaps similar outcomes could be

15. In defense of the above analysis, however, note that limiting yardstick competition may only be possible through government operation. For example, granting monopolies to private firms is politically difficult as it can easily lead to charges of corruption or favoritism. Moreover, once a monopoly is granted, future governments will have little control over the firm in case the firm turns out to do a bad job. On the other hand, if the government tries to maintain control through heavy regulation, then it might stifle the private monopoly, making it essentially government run.

achieved through a regulated private monopoly. In this subsection, we discuss another reason for lower-powered incentives with governments, reelection uncertainty faced by politicians.

To simplify the analysis in this subsection, suppose that schools can be operated either by a private monopoly or by a politician, both with the objective function

$$U_0^r = \mathbb{E}_0 \left[ \sum_{t=0}^1 (\delta^r)^t (m_t^r - C_t) \right], \quad (23)$$

where  $r$  denotes either to private monopoly or to the politician, and the only difference from equation (18) is that the discount factor is now denoted by  $\delta^r$ . Since politicians not only run schools but also control many other policies, they will be facing greater reelection uncertainty uncorrelated with their performance in running schools (i.e., uncorrelated with  $m_t^r$ ). Although the CEO of a private firm can also be fired for events unrelated to his job ability, such uncertainty is plausibly greater for politicians, who are typically entrusted to perform multiple functions.

For the analysis in this subsection, let us assume that the politician will be disqualified with probability  $1 - \pi > 0$  for reasons unrelated to his performance in schools, receiving 0 utility thereafter. This implies that the politician's effective discount factor can be written as  $\delta^{\text{pol}} \equiv \delta\pi$ . Suppose for comparison that there is no such uncertainty affecting the monopolist, so  $\delta^{\text{mon}} \equiv \delta$ . Thus, there is no difference in the preferences of politicians and monopolists, but the effective discount factor of politicians is lower because of reelection uncertainty. Suppose also that  $\beta^p = \beta$  so that without reelection uncertainty, governments would be identical to markets. Then the following proposition is immediate.

*Proposition 7.* Suppose that  $\sigma_{\theta}^2 = 0$  and  $\gamma > \underline{\gamma}$ , so that both markets and firms lead to the same inefficiently high level of effort  $g^M > g^{\text{SB}}$ , with  $g^M$  as given in Proposition 3. Then in the absence of common shocks but under reelection uncertainty, the level of effort under government operation  $g^G$  is given by

$$f'(g^G) = \frac{1}{\delta\pi\beta^p}.$$

If  $\beta^p = \beta$ , then  $g^G < g^M$  as long as  $\pi < 1$ . The level of effort with private monopoly is

$$f'(g^{\text{PM}}) = \frac{1}{\delta\beta^p}.$$

If  $\beta^p = \beta$ , then  $g^G < g^{\text{PM}} = g^M$  as long as  $\pi < 1$ .

*Proof.* See Appendix.

Higher reelection uncertainty (lower  $\pi$ ) therefore lowers the career concerns and the effort level of the politician. When there is too high-powered incentives for private monopolies, reelection uncertainty makes government operation better than private monopolies. However, there may also be too much uncertainty in the election process so that  $g^G$  drops substantially below  $g^{SB}$ . If this were the case, government operation would no longer be superior to firms or markets because now governments would be providing too low-powered incentives. This suggests that although some natural level of uncertainty in the electoral process may be helpful, if the political system is too unstable, incentives in governments may be “too weak.” Thus, government operation of certain activities is likely to be efficient only under relatively stable political regimes.

## 6. Application to Education and Pension Funds

We now discuss how our model may shed some light on why certain activities are typically operated by governments.

It is first useful to note that although some government expenditure is on typical public goods like interstate highways and scientific research, most public expenditure in developed countries is on goods that yield primarily private benefits, such as education, pensions, and health care. For example, in the United States, more than half of the noninterest, nonmilitary federal budget is spent by the Education Department, Social Security, and Health and Human Services.<sup>16</sup> In fact, governments do not simply subsidize education, savings, and health, but actually *operate* schools, pension systems, and hospitals. This is puzzling in light of the standard theories of public finance since in most cases the government can deal with market failures with Pigovian taxes and subsidies, especially given the existing evidence on widespread inefficiencies in government provision (e.g., Barberis et al. 1996; La Porta et al. 1999). Similarly, rent-seeking arguments cannot explain such government involvement because it is not clear why government should choose to be involved in education rather than in the operation of factories.<sup>17</sup>

### 6.1 Education

A large share of primary and secondary education provision is by the state in almost all countries, and in many countries this provision is highly centralized. (The United States, with its local school boards, is an exception.) Even if one accepts the case for subsidizing education, it is unclear why governments operate schools rather than simply subsidizing them. Consistent with the model, incentives appear to be weaker in government-operated schools, and there is evidence that high-powered incentives in teaching can create major

16. See US Office of Management and Budget, historical tables.

17. Bowles and Gintis (1976), Lott (1999), M. Kremer and A. Sarychev (unpublished data), and L. Pritchett (unpublished data) suggest that governments may run schools in order to control what ideology is taught to students.

distortions. Perhaps the cleanest such evidence comes from a randomized evaluation program that provided primary school teachers in rural Kenya with incentives based on students' test scores (Glewwe et al. 2003). They find that just as the model predicts, although test scores increased in treated schools, there was little evidence of more teacher effort aimed at increasing long-run learning. Teachers facing higher incentives increased effort to raise short-run test scores by conducting more test preparation sessions (i.e., "bad" type of effort). However, the "good" type of teaching did not show a proportional increase: teacher attendance did not improve, homework assignment did not increase, and pedagogy did not change. Although students in treatment schools scored higher than their counterparts in comparison schools during the life of the program, they did not retain these gains after the end of the program, consistent with the hypothesis that teachers focused more on manipulating short-run scores.

Similar results are obtained in US studies. Jacob (2002) investigates the effects of the No Child Left Behind education bill in Chicago Public Schools, which provided stronger incentives to teachers. He shows that this program led to a significant increase in math and reading achievement scores, but that these increases were influenced by teaching of test-specific skills, and that there were no comparable gains on state-administered exams. In a related study, Jacob and Levitt (2002) find substantial increases in teacher cheating (another example of "bad" type of effort) in response to the introduction of high-powered incentives in Chicago. Similarly, Figlio and Winicki (2002) look at the link between nutrition and short-term cognitive functioning, and find that school districts in Virginia increase the number of calories in school lunches on days when high-stakes tests are administered, thus artificially inflating test scores. Eberts et al. (2002), on the other hand, illustrate the potential adverse effects of a merit-based teacher incentive scheme encouraging student retention on other outcomes such as average daily attendance rates and student failure rates.

Evidence from the three countries that have moved farthest in introducing markets into education, Chile, New Zealand, and the United Kingdom, is also consistent with the notion that moving to a more market-oriented system leads to high-powered incentives and carries significant costs. Hsieh and Urqiola (2002) argue that competition among private schools in Chile's voucher program induces them to try to recruit strong students who will raise average scores and making cosmetic changes to school appearance.<sup>18</sup> Ladd and Fiske (2000) find similar effects in New Zealand. Although Glennerster (2002) has a positive overall assessment of recent British efforts to establish a quasi-market in education and publish league tables of comparative school performance, he notes that test score gains on UK exams were not matched by comparable gains on international exams. This is consistent with the possibility that schools may have focused on preparing students for the exams used to prepare the league tables, rather than on broader measures of learning.

---

18. In contrast, using an instrumental-variables strategy Gallego (2006) finds more beneficial effects of voucher schools in Chile.

## 6.2 Pension Funds

Similar issues arise in the administration of pensions. Pension systems are often run by governments, though they provide private goods. Diamond and Valdes-Prieto (1994) argue that in systems like the Chilean one, run by private firms, administrative costs are substantially higher than well-managed government-run systems. The bulk of the additional administrative costs come from “advertising,” whereby individual funds try to raise their performance appearance, and from “customers stealing,” whereby sales agents attempt to convince clients to switch from one fund to the other, without any apparent direct benefits. Both these are examples of the bad type of effort in our model. In fact, the case of pension funds is a good example of the effect of common shocks in our model. Privatizing pension funds would automatically lead to yardstick competition due to common shocks affecting the value of stocks and bonds. Thus, this industry may be particularly prone to the wasteful activities highlighted above.

In Malaysia, for example, where the government runs and manages the pension system, the Employees’ Provident Fund costs US \$10 a year per active affiliate to administer or 0.32% of annual covered earnings. In Chile, on the other hand, administrative costs average US \$51.6 a year or 1.70% of annual covered earnings. There is also evidence in Chevalier and Ellison (1999) that US mutual fund managers have significant career concerns and consequently manipulate the composition of their investments in ways that may not be in the best interest of mutual fund investors.<sup>19</sup>

The recent article by Cronqvist (2003) studies partial privatization of the Swedish pension fund system and finds that privatization led to large advertisement campaigns by private fund managers. More importantly, a bulk of this advertisement is composed of “seemingly noninformative” advertisements (another example of “bad” effort). Furthermore, the article shows that such noninformative advertisements actually lead investors “astray” by exploiting their behavioral biases. Thus even in the pension fund context, the distortionary cost of higher incentives that this article put forth can be quite large and important.

## 6.3 Health Care and Law Enforcement

Finally, our mechanism also suggests possible reasons for why health care and law enforcement may be government provided. With private provision, health care providers may compete to improve their reputation by taking actions that

---

19. The long time periods involved in pensions and the presence of relatively unsophisticated small investors make pensions more prone to signaling and potentially misleading advertising and thus increase the costs of high-powered incentives for pensions relative to other types of financial intermediation.

Moreover, though we abstract from reputation in our model, it may be particularly difficult to build reputation in the management of pension funds, in part because incentives to deviate from the high-reputation strategy would be strong. For example, a pension fund that takes big risks may have high returns in the short run and very bad outcomes only with low frequency.



make people feel better in the short run but do not improve their long-run health. For example, US hospitals provide more nonmedical amenities than British hospitals, which face less competition. Although it is certainly possible that British hospitals may be providing suboptimal nonmedical amenities, the evidence is also consistent with the notion that in the more market-based US system, hospitals are trying to signal quality by providing easily observed nonmedical amenities. This would suggest that the ratio of spending on these amenities to spending on medical care is too high in the United States, and perhaps explain why the UK manages to achieve health outcomes nearly as good as the United States, while spending only 7.3% of gross domestic product on health compared to the 13% the United States spends (OECD Health Data 2002).

Finally, law enforcement provides another example of potential cost of high-powered incentives. Law enforcement agents with too high-powered incentives may frame innocent people to appear more able to solve crimes, so regulation of incentives might also be necessary in this activity, and the low-powered incentives provided by government operation of law enforcement may be useful.

## 7. Concluding Remarks

This article has presented a model in which both high- and low-powered incentives have costs. Although high-powered incentives are necessary to induce effort from agents, they also encourage them to exert bad effort to improve observed performance. The relative importance of good and bad effort in the activity in question determines the optimal extent of incentives. The natural career concerns in market environments may then lead to too high-powered incentives. We showed how firms, envisaged as teams of producers, may be useful in this case by coarsifying the information structure and creating a moral-hazard-in-teams problem to reduce the excessively powerful incentives of agents. We also suggested that firms may sometimes be unable to do this because the naturally high-powered incentives of firm owners may trickle down to employees making it impossible to commit to low-powered incentives. In such situations, government operation might be an alternative. Governments have low-powered incentives for a variety of reasons outside our model. We also argued that there are two reasons for incentives to be low powered in governments in the context of our framework: first, government operation precludes yardsticks competition, because responsibility rests at the top; second, reelection uncertainty due to other reasons weakens politicians' incentives. Weaker politician incentives in turn may trickle down to lower-powered incentives throughout the entire government organization.

Overall, our model offers a unified framework for the analysis of the determination and implications of incentives in markets, firms, and governments. The analysis suggests that activities for which high-powered incentives are desirable should operate as markets. These would be activities where output or quality is reasonably observable and there is little scope for unproductive signaling effort. Examples of such activities may include sports, agriculture,

and simple manufacturing. As services become more complicated and there is a danger of wasteful effort due to over-incentivization, organization within firms may be appropriate where group reputation could dull incentives at the individual level. Examples may include most durable goods requiring reputation, consulting services, or journalism. Because “Mom-and-pop” operations in these fields may have too much incentive to falsely advertise and exaggerate their past performance for quality, the lower-powered incentives prevalent in large corporations might be preferable. Perhaps for this reason, durable goods retailers and many financial service firms advertise that their employees do not work on commission.<sup>20</sup>

At the other extreme, government operation may be appropriate for tasks where it is difficult for customers to accurately separate true quality from efforts to signal quality, and where firms cannot commit to low-powered incentives to build a reputation against low-quality work. This is where governments may potentially lead to better outcomes due to their ability to commit to relatively low-powered incentives to workers for reasons outlined in the article.

There are natural limits to the theory presented in this article; in practice, many other factors are important in shaping incentives in markets, firms, and governments, and the boundaries of these organizations are not simply, or perhaps even mainly, determined as a way of regulating the power of incentives. For example, governments may run certain functions for rent-seeking reasons. Nevertheless, the arguments developed in this article might suggest a reason for why government operation in some activities may be less costly than in others, thus helping us understand in which activities we are more likely to see government involvement. Overall, the importance of the forces emphasized here is therefore an empirical question. We have highlighted some existing empirical evidence regarding education and pension funds that seems to support our model. However, further empirical investigation of relative efficiency of markets, firms, and governments in different activities, taking into account issues of relative output quality and composition of effort, should be a fruitful area for future research.

## Appendix

*Proof of Lemma 1.* Although in the text we focus on the case where  $n \rightarrow \infty$ , here we solve for the general case with  $n$  finite first. Parents with prior  $m_0^i$  about teacher  $i$  observe the vector  $Z_0$ . Let  $v_0$  be the variance of  $m_0^i$ . Since  $m_0^i$  and  $Z_0$

---

20. Large corporations may also be able to build up a reputation for quality and for not encouraging their employees to mislead customers because of their high-powered incentives.

Another type of organization, which falls outside the scope of our study, is nonprofits, which may more effectively build a reputation for not allowing employees to mislead customers because they have weaker profit incentives (see Besley and Ghatak [2003], for a theory of incentives with “motivated” agents, which suggests an explanation for why certain activities should be performed in nonprofit firms).

are distributed normally, we can use the normal updating formula to compute  $m_1^i$  conditional on  $Z_0$ . In particular, we have that

$$a_1^i | Z_0 \sim N(m_1^i, v_1),$$

where

$$m_1^i = m_1^i + \Sigma_{12} \Sigma_{22}^{-1} (Z_0 - M_0) \quad (\text{A1})$$

and

$$v_1 = v_0 - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}, \quad (\text{A2})$$

with  $M_t \equiv [m_t^1 \cdots m_t^n]^T$ ,  $\Sigma_{12} \equiv [0 \ 0 \ \cdots \ v_0 \ \cdots \ 0]$  (where  $v_0$  corresponds to the  $i$ th component of the vector),  $\Sigma_{21} = [\Sigma_{12}]^T$ , and

$$\Sigma_{22} \equiv \begin{bmatrix} (v_0 + \sigma_\theta^2 + \sigma_\eta^2) & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \vdots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & (v_0 + \sigma_\theta^2 + \sigma_\eta^2) \end{bmatrix}.$$

The inverse of  $\Sigma_{12}$ ,  $\Sigma_{22}^{-1}$  can be computed as

$$\Sigma_{22}^{-1} = \frac{1}{b} \begin{bmatrix} a & 1 & 1 \\ 1 & \cdots a \cdots & 1 \\ 1 & 1 & a \end{bmatrix},$$

where

$$b \equiv (n-1)\sigma_\eta^2 - \frac{(v_0 + \sigma_\theta^2 + \sigma_\eta^2)^2}{\sigma_\eta^2} - (n-2)(v_0 + \sigma_\theta^2 + \sigma_\eta^2) \quad \text{and}$$

$$a \equiv -\left[ \frac{(v_0 + \sigma_\theta^2 + \sigma_\eta^2)}{\sigma_\eta^2} + (n-2) \right].$$

Plugging in the value of  $\Sigma_{12}$  and  $\Sigma_{22}^{-1}$ , we obtain

$$m_1^i = m_0^i + \beta(z_0^i - m_0^i) - \bar{\beta}(\bar{z}_0^{-i} - \bar{m}_0^{-i}), \quad (\text{A3})$$

where

$$\bar{z}_0^{-i} = \frac{1}{(n-1)} \sum_{k \neq i} z_0^k$$

$$\bar{m}_0^{-i} = \frac{1}{(n-1)} \sum_{k \neq i} m_0^k$$

$$\beta = \frac{v_0 a}{b} \quad \text{and} \quad \bar{\beta} = -\frac{v_0(n-1)}{b}.$$

Note that  $1 > \beta \geq \bar{\beta} > 0$ .

We can also solve for  $v_1$  from the above equation (A2) as

$$v_1 = v_0 - \frac{a}{b}v_0^2.$$

As  $n \rightarrow \infty$ , we obtain the expressions in the text:

$$\beta = \frac{v_0}{(v_0 + \sigma_\theta^2)} \quad \text{and} \quad \bar{\beta} = \frac{v_0}{(v_0 + \sigma_\theta^2)},$$

as well as

$$v_1 = \frac{v_0\sigma_\theta^2}{v_0 + \sigma_\theta^2},$$

for all  $i$ .

Note that as  $n \rightarrow \infty$ , the expression for  $\beta$  is the same as if the parents initially knew the exact value of  $\eta_0$ . This is because as  $n \rightarrow \infty$  parents can perfectly filter out the common shock  $\eta_0$  by averaging  $z_0^i - m_0^i$  across different teachers.

This argument holds even if  $v_0^i$  depends on  $i$ . By the standard normal updating rule applied to this case, we have that

$$m_1^i = m_0^i + \beta_i(z_0^i - \hat{\eta}),$$

$$\text{where } \beta_i = \frac{v_0^i}{(v_0^i + \sigma_\theta^2)}. \quad \blacksquare$$

*Proof of Proposition 2.* The second-best is given by the solution to equation (6)

$$\delta f'(g_0) \frac{\partial \mathbb{E}_0[w_1(\Omega_1^i)]}{\partial s_0^i} = 1 \quad \text{and} \quad \gamma \delta f'(b_0) \frac{\partial \mathbb{E}_0[w_1(\Omega_1^i)]}{\partial s_0^i} = 1. \quad (\text{A4})$$

The above conditions can be combined to give  $f'(g_0) = \gamma f'(b_0)$  which implies that:

$$b_0 = f'^{-1}\left(\frac{f'(g_0)}{\gamma}\right). \quad (\text{A5})$$

The inverse of  $f'(x)$  exists due to the concavity of  $f(x)$ . Equation (A5) defines the *feasible* pairs of  $(g_0, b_0)$  even when the wage function is not differentiable. Restriction to differentiable wage functions therefore does not change our second-best solution. Given (A5), we can simplify our maximization problem (6) into the unconstrained problem:

$$\max_{g_0} \left( \bar{A} + f(g_0) - g_0 - f'^{-1}\left(\frac{f'(g_0)}{\gamma}\right) \right). \quad (\text{A6})$$

The above is a well-defined maximization problem with a unique global maximum. Differentiating (A6) with respect to  $g_0$ , we obtain

$$f'(g_0^{\text{SB}}) = 1 + \frac{(1/\gamma)f''(g_0^{\text{SB}})}{f''(f'^{-1}(f'(g_0^{\text{SB}})/\gamma))}.$$

Because of the concavity of  $f(x)$ , the expression on the right-hand side is greater than 1, implying  $g_0^{\text{SB}} < g_0^{\text{FB}}$ . The value of  $b_0^{\text{SB}}$  is determined by  $g^{\text{SB}}$  and equation (A5).

Given  $g^{\text{SB}}$ , any wage structure  $w_1(\Omega_1^i)$  that implements  $g^{\text{SB}}$  will also implement  $b_0^{\text{SB}}$ , given that the choice implemented by any wage structure satisfies (A4). Since the decision  $g^{\text{SB}}$  can always be implemented by some wage schedule of the form  $w_1^i = \alpha m_1^i + \kappa$ , there is no loss of generality to restrict ourselves to such linear schedules. To see this, note that from equation (3),

$$m_1^i = (1 - \beta)m_0^i + \beta s_0^i + \text{constant}.$$

We can therefore write

$$\mathbb{E}_0[w_1^i] = \alpha(1 - \beta) \cdot \mathbb{E}_0[m_0^i] + \alpha\beta \cdot s_0^i + \kappa,$$

what implies that  $\frac{\partial \mathbb{E}_0[w_1(\Omega_1^i)]}{\partial s_0} = \alpha\beta$ . By choosing  $\alpha = \alpha^{\text{SB}}$ , where

$$\alpha^{\text{SB}} = \frac{1}{f'(g^{\text{SB}})\beta\delta},$$

one can always implement  $g^{\text{SB}}$ . Finally, since  $\alpha^{\text{SB}}$  is an increasing function of  $g^{\text{SB}}$ , which is decreasing in  $\gamma$ , there exists a  $\underline{\gamma}$  such that for  $\gamma > \underline{\gamma}$ , we have  $\alpha^{\text{SB}} < 1$ . ■

*Proof of Proposition 4.* With firms, parents observe  $J$  signals, represented by

$$\bar{Z}_0 \equiv [\bar{z}_0^1 \bar{z}_0^2 \dots \bar{z}_0^J]^T,$$

where

$$\bar{z}_0^j = \bar{a}_0^j + \bar{\theta}_0^j + \eta_0.$$

Once more, the Bayesian updating of beliefs will imply that given the prior

$$\bar{a}_0^j \sim \mathcal{N}(\bar{m}_0^j, v_0/K_j),$$

and the fact that  $\bar{Z}_0$  is normally distributed, we will have that

$$\bar{a}_1^j | \bar{Z}_0 \sim \mathcal{N}(\bar{m}_1^j, v_1^j).$$

By hypothesis, there are  $J$  firms in the economy with each firm  $j$  having a size  $K_j$ . Let  $n \rightarrow \infty$ . Since the size of each school is finite, as  $n \rightarrow \infty$  we also have that  $J \rightarrow \infty$ , that is there are infinitely many firms. Therefore, parents can obtain a consistent estimator of  $\eta_0$  by averaging  $\bar{z}_0^j - \bar{m}_0^j$  across different schools. Since parents are Bayesian players that can empirically estimate arbitrarily well  $\eta_0$  with some estimator  $\hat{\eta}$ , while updating their beliefs about each individual school, they will estimate the posterior distribution of  $\bar{a}_0^j$  as if they

knew  $\eta_0$  and  $\eta_0 = \hat{\eta}$ , for some estimator  $\hat{\eta}$ . Moreover, the observation of one single school  $\bar{z}_0^j$ , where  $j$  is fixed, should not matter for  $\hat{\eta}$ . Those points were already noticed in the proof of Lemma 1.

Finally, the observations on other schools are only useful for parents estimating  $\bar{a}_0^j$  as long as it provides information on  $\eta_0$ . These facts together imply that, as  $n \rightarrow \infty$ , estimating  $\bar{m}_1^j$  should be the same estimating  $\bar{m}_1^j$  when parents observe only

$$\bar{z}_0^j | \bar{a}_0^j \sim \mathcal{N}(\bar{a}_0^j, \sigma_0^2/K_j + \sigma_{\hat{\eta}}^2/K_j),$$

and know that  $\eta_0 = \hat{\eta}$ , where  $\hat{\eta}$  is a function of  $\bar{Z}_0$  that does not depend on  $\bar{a}_0^j$ .

The Bayesian updating rule for this simple case is given by

$$\bar{m}_1^j = \bar{m}_0^j + \beta \cdot (\bar{z}_0^j - \hat{\eta} - \bar{m}_0^j),$$

where  $\beta = \frac{v_0}{v_0 + \sigma_0}$ .

In other words, as  $J \rightarrow \infty$ , the career concerns coefficient for the entire firm is exactly the same as the career concerns coefficient for an individual teacher under market equilibrium. However, the career concerns coefficient for an *individual*  $i$  in firm  $j$  is given by  $\beta/K_j$ , and is decreasing in  $K_j$ .

From this, it is straightforward to see that  $g^F(1) = g^M$  and  $g^F(K_j) \rightarrow 0$  as  $K_j \rightarrow \infty$ . Moreover,  $g^F(K_j)$  is monotonically decreasing in  $K_j$ . The firm will now endogenously set  $K_j = K^*$  such that  $g^F(K^*) = g^{SB}$ . To see this, note that the firm will maximize

$$\max_{K_j} \mathbb{E} \left[ \sum_{t=0}^1 \delta^t (\bar{m}_t^j + f(g_t(K_j)) - g_t(K_j) - b_t(K_j)) \right]. \tag{A7}$$

Since there is no effort in period 1, the above problem is the same as in equation (6) in Proposition 2, and is maximized at  $K_j = K^*$ , such that  $g(K^*) = g^{SB}$ , providing the second-best solution. ■

*Proof of Proposition 5.* Let  $n \rightarrow \infty$  so that with each firm of finite size,  $J \rightarrow \infty$ . As before, this assumption implies that the common shocks can be perfectly filtered out, so to simplify notation, we ignore the common shocks. Since there is asymmetric information now, we must distinguish between internal and public information. The internal information on an individual  $i$  in firm  $j$  can be summarized by

$$z_t^{ji} = a_t^{ji} + \theta_t^{ji} + \tilde{\theta}_t^{ji},$$

whereas the public information is given by  $\bar{z}_t^j = \bar{a}_t + \bar{\theta}_t^j$ . The firm has access to both internal and public information. Recall that each teacher also gets her full surplus, that is,  $w_t^{ji} = m_t^{ji} + f(g_t)^j$ .

Given the internal and public signal, the updating formula used by the firm becomes

$$m_{t+1}^{ji} = m_t^{ji} + \left[ v_0 \quad \frac{v_0}{K} \right] \begin{bmatrix} (v_0 + \sigma_\theta^2 + \sigma_\xi^2) & \frac{(v_0 + \sigma_\theta^2)}{K} \\ \frac{(v_0 + \sigma_\theta^2)}{K} & \frac{(v_0 + \sigma_\theta^2)}{K} \end{bmatrix}^{-1} \begin{bmatrix} (z_t^{ji} - m_t^{ji}) \\ (\bar{z}_t^j - \bar{m}_t^j) \end{bmatrix},$$

which implies

$$m_{t+1}^{ji} = m_t^{ji} + \beta^{\text{asy}}(z_t^{ji} - m_t^{ji}) + \bar{\beta}^{\text{asy}}(\bar{z}_t^{j(-i)} - \bar{m}_t^{j(-i)}),$$

where

$$\beta^{\text{asy}} \equiv \left( \frac{v_0(K-1)(v_0 + \sigma_\theta^2) + v_0\sigma_\theta^2}{(v_0 + \sigma_\theta^2)((K-1)(v_0 + \sigma_\theta^2) + K\sigma_\theta^2)} \right)$$

defines the career concerns coefficient with asymmetric information, super-script  $-i$  refers to the average excluding the  $i$ th teacher, and

$$\bar{\beta}^{\text{asy}} \equiv \frac{v_0\sigma_\theta^2(K-1)}{(v_0 + \sigma_\theta^2)((K-1)(v_0 + \sigma_\theta^2) + K\sigma_\theta^2)(K-1)}.$$

To emphasize dependence on firm size, let us write the career concerns coefficient above,  $\beta^{\text{asy}}$ , as  $\beta^{\text{asy}}(K)$ . Let  $K^{**}$  be the value of  $K$  that makes  $\beta^{\text{asy}}(K) = \beta_{\text{SB}}$ . Then, we have that  $\frac{\partial K^{**}}{\partial \sigma_\theta^2} < 0$ . In other words, as the firm learns more about an individual teacher, it becomes harder to sustain the second-best level of effort, and firm size needs to increase.

Since  $\frac{\partial \beta^{\text{asy}}(K)}{\partial K} < 0$ , and  $\beta^{\text{asy}}(K=1) = \beta_M$ , to establish  $\beta^{\text{asy}}(K) = \beta_{\text{SB}}$ , we simply need to show that  $\lim_{K \rightarrow \infty} \beta^{\text{asy}}(K) < \beta_{\text{SB}}$ . We have

$$\lim_{K \rightarrow \infty} \beta^{\text{asy}}(K) = \frac{v_0}{v_0 + \sigma_\theta^2 + \sigma_\xi^2}$$

and

$$\left( \frac{v_0}{v_0 + \sigma_\theta^2 + \sigma_\xi^2} \right) < \beta_{\text{SB}} \Leftrightarrow \sigma_\theta^2 > \bar{\sigma}_\theta^2 \equiv \left( \frac{v_0}{\beta_{\text{SB}}} - v_0 - \sigma_\theta^2 \right).$$

Therefore, if  $\sigma_\theta^2 > \bar{\sigma}_\theta^2$ , the economy can achieve the second-best allocation. However, for  $\sigma_\theta^2 \leq \bar{\sigma}_\theta^2$  (i.e., severe asymmetric information), the commitment problem implies that the second-best can never be achieved. ■

*Proof of Proposition 6.* The politician's maximization problem is

$$\max_{\alpha_1^p} U_0^{\text{pol}} = \mathbb{E}_0 \left[ \sum_{t=0}^1 \delta^t (m_t^p - \bar{w}_t) \right].$$

Notice that  $m_0^p$  is given, and wage  $\bar{w}_0$  is paid up front in the first period. Given participation constraint  $\underline{u}$  for a teacher,  $\bar{w}_0 = \underline{u} + g_0 + b_0$ , where  $g_0$  and  $b_0$  are

effort levels induced by the government for the average teacher. These effort levels in the first period are induced by giving an incentive to each teacher of  $w_1^i = \alpha_1^p m_1^i + \kappa_1$  in the second period. Since teachers do not exert any effort in the second period, there is no need to compensate them for any effort in second period. Furthermore, since in expected terms there is no uncertainty about average ability of the teachers in second period, second period wage drops out of the maximization problem.

Thus, the above maximization problem for the politician reduces to  $\max_{\alpha_1^p} U_0^{\text{pol}} = \mathbb{E}_0[\delta m_1^p - g_0 - b_0]$ , where  $m_1^p = m_0^p + \beta^p(z_0 - \lambda m_0^p)$ . Replacing  $m_1^p$  with its updating equation, and substituting for  $z_0$ , the politician's maximization problems can be rewritten as

$$\max_{\alpha_1^p} \mathbb{E}_0[\delta \beta^p (f(g_0) + \gamma f(b_0)) - g_0 - b_0],$$

where

$$\{g_0, b_0\} \in \arg \max_{\{g_0, b_0\}} (\delta \alpha_1^p m_1^i - g_0' - b_0').$$

The politician's first order condition with respect to  $\alpha_1^p$  gives us:

$$\delta \beta^p \left( f'(g_0) \frac{\partial g_0}{\partial \alpha^p} + \gamma f'(b_0) \frac{\partial b_0}{\partial \alpha^p} \right) = \left( \frac{\partial g_0}{\partial \alpha^p} + \frac{\partial b_0}{\partial \alpha^p} \right).$$

Each teacher's first order condition gives us  $f'(g_0) = \gamma f'(b_0)$ , and plugging this into the politician's FOC, we get the final expression

$$f'(g^G) = \frac{1}{\delta \beta^p}. \quad \blacksquare$$

*Proof of Proposition 7.* The proof is identical to that of Proposition 6, but with  $\beta^p = \beta$ , and  $\delta$  replaced by  $\pi \delta$ . ■

## References

- Acemoglu, Daron. 1996. "Career Concerns and Coverups: A Dynamic Model of Collusion." MIT Working Paper.
- . 1998. "Credit Market Imperfections and Separation of Ownership from Control," 78 *Journal of Economic Theory* 355–81.
- Acemoglu, Daron, Michael Kremer, and Atif Mian. 2003. "Incentives in Markets, Firms and Governments." NBER Working Paper No. 9802.
- Acemoglu, Daron, and Thierry Verdier. 2000. "The Choice Between Market Failures and Corruption," 90 *American Economic Review* 194–211.
- Acuna, Rodrigo, and Augusto Iglesias. 2001. "Chile's Pension After 20 Years." World Bank Social Protection Discussion Paper Series No. 0129.
- Banerjee, Abhijit. 1997. "A Theory of Misgovernance," 112 *Quarterly Journal of Economics* 1289–332.



- Barberis, Nicholas, Maxim Boycko, Andrei Shleifer, and Natalia Tsukanova. 1996. "How does Privatization Work? Evidence from the Russian Shops," 104 *Journal of Political Economy* 764–90.
- Bates, Robert. 1981. *Markets and States in Tropical Africa*. Berkeley, Calif.: University of California Press.
- Besley, Timothy, and Maitreesh Ghatak. 2003. "Competition and Incentives with Motivated Agents." London School of Economics Mimeo.
- Bowles, Samuel, and Herbert Gintis. 1976. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. New York, N.Y.: Basic Books.
- Chevalier, Judith, and Glenn Ellison. 1999. "Career Concerns of Mutual Fund Managers," 114 *Quarterly Journal of Economics* 389–431.
- Coase, Robert. 1937. "The Nature of the Firm," 4 *Economica* 386–405.
- Cronqvist, Henrik. 2003. "Advertising and Portfolio Choice." University of Chicago, GSB Mimeo.
- Crozier, Maurice. 1967. *Bureaucratic Phenomenon*. Chicago: University of Chicago Press.
- Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures," 66 *Review of Economic Studies* 183–98.
- Diamond, Peter, and Salvador Valdes-Prieto. 1994. "Social Security Reforms," in Bosworth, Barry P., Rudiger Dornbusch, and Raul Laban, eds., *The Chilean Economy: Policy Lessons and Challenges*. Washington, D.C.: The Brookings Institution, 257–328.
- Dixit, Avinash. 1997. "Power of incentives in private vs. public organizations," 87 *American Economic Review, Papers and Proceedings* 378–82.
- . 2002. "Incentives and organizations in the public sector: An interpretative review," 37 *Journal of Human Resources* 696–727.
- Eberts, Randall, Hollenbeck, Kevin, and Stone, Joe. 2002. "Teacher Performance Incentives and Student Outcomes," 37 *Journal of Human Resources*. 913–27.
- Figlio, David N., and Joshua Winicki. 2002. "Food for Thought: The Effects of School Accountability Plans on School Nutrition." National Bureau of Economic Research Working Paper 9319.
- Gallego, Francisco. 2006. "Voucher-School Competition, Incentives and Outcomes: Evidence from Chile." MIT Mimeo.
- Gibbons, Bob. 1998. "Incentives in Organizations." National Bureau of Economic Research Working Paper 6695.
- Gibbons, Bob, and Murphy, Kevin. 1999. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," 100 *Journal of Political Economy* 468–505.
- Glennerster, Howard. 2002. "United Kingdom Education 1997–2001," 18 *Oxford Review of Economic Policy* 120–36.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2003. "Teacher Incentives." National Bureau of Economic Research Working Paper 9671.
- Grossman, S., and O. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," 94 *Journal of Political Economy* 691–719.
- Hannaway, Jane. 1992. "Higher Order Skills, Job Design, and Incentives: An Analysis and Proposal." *American Educational Research Journal* 29(1): 3–21.
- Hanushek, Eric A. 1996. "Outcomes, Cost, and Incentives in Schools," in Hanushek, Eric A. and Dale W. Jorgenson, eds., *Improving America's Schools: The Role of Incentives*. Washington, D.C.: National Academies Press 29–52.
- Hart, O., and J. Moore. 1990. "Property Rights and the Nature of the Firm," 98 *Journal of Political Economy* 1119–58.
- Hart, O., A. Shleifer, and R. Vishny. 1997. "The Proper Scope of Government: Theory and an Application to Prisons," 112 *Quarterly Journal of Economics* 1127–61.
- . 1999. "Managerial Incentive Problems: A Dynamic Perspective," 66 *Review of Economic Studies* 169–82.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," 7 *Journal of Law, Economics and Organization* 24–52.
- . 1994. "The Firm as an Incentive Structure," 84 *American Economic Review* 972–91.

- Hieh, Chang-Tai and M. Urquiola. "The Effects of Generalized School Choice on Achievement: Evidence from Chile's Voucher Program." 90 *Journal of Public Economics* 1477–1503.
- Jacob, B. 2002. "Accountability, Incentives, and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools." National Bureau of Economic Research Working Paper 8968.
- Jacob, B., and S. Levitt. 2002. "Teacher Cheating." University of Chicago Mimeo.
- Johnson, Ronald N., and Gary D. Libecap. 1994. *The Federal Civil Service and the Problem of Bureaucracy: The Economics and Politics of Institutional Change*. Chicago: University of Chicago Press.
- Kihlstrom, R., and M. Riordan. 1984. "Advertising as a Signal," 92 *Journal of Political Economy* 427–50.
- Kremer, Michael. 1997. "Why Are Worker Cooperatives So Rare?" National Bureau of Economic Research Working Paper 6118.
- Ladd, Helen, and Edward Fiske. 2000. *When Schools Compete: A Cautionary Tale*. Washington, D.C.: The Brookings Institution.
- La Porta R., F. López-de-Silanes, A. Shleifer and R. Vishnyt. 1999. "The Quality of Government," *Journal of Law Economics and Organization*.
- Levin, J., and S. Tadelis. 2002. "A Theory of Partnerships." Stanford University Working Paper.
- Lott, John, Jr. 1999. "Public Schooling, Indoctrination and Totalitarianism," 107 *Journal of Political Economy* 127–57.
- Meyer, M., and J. Vickers. 1997. "Performance Comparisons and Dynamic Incentives," 105 *Journal of Political Economy* 547–81.
- Martimort, D. 1996. "The Life Cycle of Regulatory Agencies: Dynamic Capture and Transaction Costs," 66 *Review of Economic Studies* 929–47.
- Milgrom, P., and J. Roberts. 1986. "Price and Advertising Signals of Product Quality," 94 *Journal of Political Economy* 796–821.
- Niskanen, William A. 1971. *Bureaucracy and Representative Government*. Chicago, Ill.: Aldine/Atherton.
- OECD Health Data. 2002. *A Comparative Analysis of 30 Countries*. CD-Rom. Paris: OECD.
- Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency," 111 *Journal of Political Economy* 929–58.
- Shleifer, Andrei, and Robert W. Vishny. 1994. "Politicians and Firms," 109 *Quarterly Journal of Economics* 995–1025.
- Stein, J. 1989. "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," 104 *Quarterly Journal of Economics* 655–69.
- Tirole, Jean. 1986. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," 2 *Journal of Law Economics and Organization* 181–214.
- Williamson, O. 1985. *The Economic Institutions of Capitalism*. New York, N.Y.: Free Press.
- Wilson, James. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York: Basic Books Classics.