Incentives in Markets, Firms and Governments*

Daron Acemoglu
Department of Economics, Massachusetts Institute of Technology

Michael Kremer
Department of Economics, Harvard University

Atif Mian
Graduate School of Business, University of Chicago†

July 2004

Abstract

We construct a simple career concerns model where high-powered incentives can distort the composition of effort by stimulating unproductive signalling. In such a setting, markets 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 may have a commitment problem if the owner has insider information on individual employees. We show that in these circumstances governments may turn out to be the only organizational form able to credibly commit to low-powered incentives even if run by a self-interested politician. Among other reasons, this may happen because of the government's ability to limit yardstick competition and re-election uncertainty. We discuss possible applications of our theory to pervasive government involvement in predominantly private goods such as education and management of pension funds.

Keywords: career concerns, incentives, moral hazard in teams, signaling, multi-tasking, schools, team production, teachers.

JEL Classification: D23, L22, H10, H52.

---

*We thank Avinash Dixit, Mathias Dewatripont, Robert Gibbons, Joseph Harrington, Bengt Holmstrom, Canice Prendergast, Raghuram Rajan, Jean Tirole, Jan Zabojnik, two anonymous referees, and seminar participants at the NBER economics organization conference for useful comments and Jean Lee, Janina Matuszeski, and Courtney Umberger for research assistance.

†Email: daron@mit.edu, mkremer@fas.havard.edu, atif@gsb.uchicago.edu.

1 Introduction

While a range of transactions take place in markets and are subject to strong incentives, many important activities are organized within firms that are partly shielded from market incentives. Still others are conducted by governments, where they are even more insulated from market incentives. While the costs of flat incentives are well-known, a body of work beginning with Holmstrom and Milgrom [1991] suggests that in some circumstances weak incentives may actually be optimal. In this paper, 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. Following Holmstrom and Milgrom [1991], 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 distortion towards unproductive effort with high-powered incentives is sufficiently severe, low-powered incentives may be optimal. However we show that due to competitive pressures, individuals competing in markets will face high-powered incentives. Firms, on the other hand, may be able to "coarsify" information by organizing activity (e.g. teaching) into teams, thus providing low-powered incentives. However, since firms themselves compete in the market, if firm owners have inside information about individual team members (teachers) and can reward them secretly, firms may not be able to credibly commit to low-powered incentives.

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, for example, Dixit [1997, 2002]) about why governments can better commit to low-powered incentives. First, if the ability or actions of the politician do not matter for student performance (other than through the incentives provided to teachers), then there is no commitment problem in promising to provide civil servants with low-powered incentives. Second, in the presence of common shocks, even if the ability and actions of the politician matter, and politicians are driven by self-interest, politicians will still be able to commit to lower incentives because of the weakening of yardstick competition associated with government provision. Third, politicians may be less subject to career concerns incentives because of the re-election uncertainty that they face.

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 towards the "bad" type of effort is particularly high may be organized within governments (see section 6 for more details).[1] This adds to 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 e.g., Niskanen [1971], Bates [1981], 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 paper does not claim to offer a complete theory of the division of economic activities between markets, firms and governments—many other factors influence the boundaries of firms, with ownership of assets being a dominant explanation (see, among others, Williamson [1985], Grossman and Hart [1986], and Hart and Moore [1990]). Nevertheless, we hope that our model will be complementary to existing theories of organization, a particular, as it provides a simple unified framework for thinking about markets, firms, and governments. While traditional literature has focused on how asset ownership shapes investment incentives, our approach is based on the ability of an organizational structure to credibly commit to manipulating information.[2]

[1] Wilson [1989] and Dixit [2002] emphasize the importance of multiple tasks in the public sector.

[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 specialize in areas where they have a comparative advantage while also growing beyond this set of activities, and thus beyond their optimal size, because of rent-seeking or other political economy reasons.

This paper is closely related to the career concerns literature (e.g., Holmstrom [1999], Stein [1989], Meyer and Vickers [1997], Dewatripont, et al. [1999]), and to the multi–tasking literature (e.g., Holmstrom and Milgrom [1991] and [1994]), which also emphasizes the costs of high-powered incentives.[3] 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 other authors, in particular Gibbons [1998] and Gibbons and Murphy [1992], 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. Papers by Kremer [1997] and Levin and Tadelis [2002] are also related—they emphasize the benefits of firms in manipulating incentives because of joint production, though their story is non-informational and static.

Perhaps most closely related to our paper is the contribution by Hart, Shleifer and Vishny [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 paper 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.

The rest of the paper is organized as follows. Section 2 describes the environment and characterizes optimal incentives in a simple mechanism design problem. Sections 3, 4, and 5 then 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.

[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]).

2 Model

We now outline the basic model. For concreteness, we will focus on the teaching example.

2.1 The environment

Consider an infinite horizon economy with n infinitely lived teachers, and $n' > n$ parents in every period, each with one child to be educated. $K = 1, 2, ...$ children can be taught jointly by K teachers. Each teacher, i, is endowed with a teaching ability a_t^i at the beginning of period t. The level of a_t^i is unknown, but both teacher i and parents share the same belief about the distribution of a_t^i. The common belief about teacher i's ability at time t is:

$$a_t^i \backsim N(m_t^i, v_t).$$

Ability evolves over time according to the stochastic process given by:

$$a_{t+1}^i = a_t^i + \varepsilon_t^i, \tag{1}$$

where ε_t^i is i.i.d. with $\varepsilon \backsim N(0, \sigma_\varepsilon^2)$. The disturbance term ε could result from personal shocks, or it could reflect the ability of the teacher to adapt to changing education demands and technology.

We consider a multi-tasking environment where a teacher can exert two types of effort, "good" and "bad", denoted by g_t^i and b_t^i respectively. The titles "good" and "bad" reflect the social value of these efforts. The human capital, h_t^j of child j is given by:

$$h_t^j = \overline{a}_t^j + \overline{f(g_t)}^j \tag{2}$$

where $\overline{a}_t^j = \frac{1}{K^j} \sum\limits_{i \in \mathbb{K}^j} a_t^i$ and $\overline{f(g_t)}^j = \frac{1}{K^j} \sum\limits_{i \in \mathbb{K}^j} f(g_t^i)$ with $\mathbb{K}^j$ is the set of teachers teaching child j, and K^j as the number of teachers in the set $\mathbb{K}^j$. In addition, $f(g)$ is increasing and strictly concave in g, with $f(0) = 0$, and $h_t^i = 0$ if the child is not taught by a teacher.

In this section, it is useful to start with the case where each child is taught by a single teacher, in which case (2) specializes to

$$h_t^i = a_t^i + f(g_t^i), \tag{3}$$

where, in this case, we can index the child taught by teacher i by i.

Parents only care about the level of human capital provided to their children. The expected utility of a parent at time t is given by:

$$U_t^P = E_t\left[h_t\right] - w_t,$$

where $E_t\left[\cdot\right]$ denotes expectations with respect to publicly available information at the beginning of time t and $\delta < 1$ is the discount rate, and w is the wage paid to the teacher.

The expected utility of a teacher i at time t is given by the time separable utility function:

$$U_t^i = E_t\left[\sum_{\tau=0}^{\infty}\delta^\tau\left(w_{t+\tau}^i - g_{t+\tau}^i - b_{t+\tau}^i\right)\right],$$

where $w_{t+\tau}^i$ denotes the wage of the teacher at time $t+\tau$.

The level of h_t^i provided by a teacher is not observable to parents. Instead, parents have to rely on an imperfect signal of h, given by the test scores, s. The test score of child j in the general case is given by:

$$s_t^j = h_t^j + \gamma\overline{f(b_t)}^j + \overline{\theta}_t^j + \eta_t, \tag{4}$$

where $\gamma \geq 0$, θ_t^i is an i.i.d. student-level shock distributed as $N(0, \sigma_\theta^2)$, for example, the ability of the students to learn, and η_t is a common shock that *every* teacher receives in period t. For example, if all students are given the same test, η_t can be thought of as the overall difficulty of the test, or any other cohort-specific difference in ability or the curriculum. η_t is distributed i.i.d. and $N(0, \sigma_\eta^2)$. In addition, $\overline{f(b_t)}^j$ and $\overline{\theta}_t^j$ are defined analogously as averages over the set of teachers in $\mathbb{K}^j$. In the special case of this section where each child is taught by a single teacher, we have:

$$s_t^i = h_t^i + \gamma f(b_t^i) + \theta_t^i + \eta_t. \tag{5}$$

Notice that the variance σ_θ^2 measures the quality of signal s_t^i, while the variance of the common shock, σ_η^2, also affects the informativeness of the signal. The lower these variances, the more precise the signal is in measuring the human capital contribution of teachers. Notice that the signal of human capital is imperfect in two ways. First, shocks θ and η make the test score a noisy signal for the student's human capital. Second, the signal can be inflated by bad effort. The parameter γ measures the extent to which the signal can be manipulated by bad effort. It also captures the importance of output quality and composition of effort relative to the amount of effort (i.e., as γ declines, the importance of ensuring a high level of total effort increases).

The reason for calling the two types of efforts good and bad should be apparent now. Parents care about the good effort exerted by the teacher, but only observe the signal s, which can be manipulated by bad effort. In practice, bad effort may correspond to what is commonly referred to as "teaching to the test". It involves rote learning, where the teacher just forces the students to cram certain essential facts or methods, without explaining the concepts behind them or the connection between the various facts and phenomena (see Hanaway [1992]). Such cramming is less useful than good effort in terms of the human capital of the students, 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. In the context of the pension funds example as well as some other applications, we can think of bad effort as any activity that improves observed performance without affecting actual performance equally (e.g., advertising).

The timing of events in this world is as follows. In the beginning of every period t, parents form priors, m_t^i, on the abilities of teachers based on the histories of test scores of the teachers. They then offer a wage w_t^i based on the expected ability of the teacher working with their child. The teacher then decides on the levels of good and bad effort, and h and s are realized at the end of period t. Ability a_t^i is then updated according to the stochastic process (1). The process then repeats itself in period $t + 1$.

We characterize the rational expectations equilibrium path where all teachers choose $\left\{g_{t+\tau}^i, b_{t+\tau}^i\right\}_{\tau=0,1,..}$ optimally given their rewards, and the beliefs about teacher ability are given by Bayesian updating. We also focus on the long-run of the model so that the variance of each teacher's ability is constant, i.e. $v_t = v_{t+1} = v$. Finally, in the text, we focus on the case where n is very large, i.e., $n \to \infty$ (the Appendix gives equations that apply for $n < \infty$ as well). The $n \to \infty$ assumption allows us to ignore the common shock, η_t, which can be backed out from the average of all test score signals in the population. The common shock will play an important role in Section 5 when we discuss the incentives with government-provided education.

2.2 Updating beliefs

Parents' belief about teacher i at the beginning of period t can be summarized as, $a_t^i \backsim N(m_t^i,\, v_t)$. Let $S_t = [s_t^1 \text{ ...... } s_t^n]^T$ denote the vector of n test scores that the agents observe during period t when each child is taught by a single teacher, i.e., equations (3) and (5). Along the rational expectations equilibrium path, parents correctly infer effort levels g_t^i and b_t^i chosen by the teachers. This means that parents can back out the part of S_t which only reflects the ability levels of the teachers, plus the noise. Let $Z_t = [z_t^1 \text{ ...... } z_t^n]^T$ denote this backed out signal, where

$$\begin{aligned} z_t^i &= s_t^i - f(g_t^i) - \gamma f(b_t^i) \\ &= a_t^i + \theta_t^i + \eta_t \end{aligned}$$

Let a_{t+1}^i be the updated prior on teacher i's ability conditional on observing Z_t. Then the normality of the error terms and the additive structure in equation (5) imply that $a_{t+1}^i \backsim N(m_{t+1}^i,\, v_{t+1})$ where m_{t+1}^i and v_{t+1} denote the mean and the variance of the posterior distribution. Using the normal updating formula, setting $v_{t+1} = v_t = v$ and focusing on the limit $n \to \infty$, we obtain the law of motion of the posterior of teacher i's mean ability, m_{t+1}^i, as:

$$m_{t+1}^i = m_t^i + \beta(z_t^i - m_t^i) - \overline{\beta}(\overline{z}_t^{-i} - \overline{m}_t^{-i}), \tag{6}$$

where

$$\beta = \overline{\beta} = \frac{1 + \sqrt{1 + 4\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right)}}{1 + 2\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right) + \sqrt{1 + 4\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right)}}, \tag{7}$$

z_t^i is the ith element of the vector Z_t, and refers to the signal from teacher i, while $\overline{z}_t^{-i}$ is the average test score excluding teacher i. Since $n \to \infty$, we have $(\overline{z}_t^{-i} - \overline{m}_t^{-i}) \to \eta_t$, so the common shock is revealed and filtered out. The proof of (6) and (7) is given in the Appendix, where we also provide the expressions for the case of n finite.

The equation (6) illustrates the relative performance evaluation (yardstick competition) in the presence of η_t. The coefficient $\overline{\beta}$ captures relative performance evaluation. It emphasizes that an improvement in the score of a teacher creates a negative effect on the market's assessment of other teachers.

Lemma 1 Parents update their beliefs about teacher $i's$ ability level, according to equations (6) and (7), where $1 > \beta > 0$. β is increasing in σ_ε^2, and decreasing in σ_θ^2.

The intuition behind the last part of Lemma 1 is that any increase in the variance of θ, σ_θ^2, increases the noise in the signal, and makes it less valuable, and hence reduces β. An increase in σ_ε^2 makes the signal more valuable due to a greater change in ability since last period. In other words, a greater σ_θ^2 relative to σ_ε^2 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

We define social welfare at time t, U_t^W, as the sum of the teachers' and parents' utilities. Since the ability of teacher i enters additively in their utility function, all teachers should choose the same effort level in a given period. Social welfare can then be written as:

$$U_t^W = \sum_{\tau=0}^{\infty} \delta^\tau (\overline{A} + f(g_{t+\tau}) - g_{t+\tau} - b_{t+\tau}) \tag{8}$$

where $\overline{A}$ is the average ability of teachers in the population, which is constant when $n \to \infty$, and $g_{t+\tau}$ and $b_{t+\tau}$ are the good and bad effort levels chosen by all teachers.

First Best: Maximizing (8) gives us the first-best. In the first-best, there is no bad effort, $b_t = 0$, and the level of good effort, g^{FB}, is given by $f'(g^{FB}) = 1$.

Second-Best: Since teacher effort and the level of human capital are not directly observable, a more useful benchmark is given by solving for the optimal mechanism given these informational constraints.

Let $\Omega_t^i = [m_0^i \ s_0^i \ s_1^i \ s_2^i \ ...... \ s_{t-1}^i]$ be the information set containing the vector of test scores for teacher i at the beginning of period t when all children are taught by a single teacher. Ω_t^i is the largest set of contractible information about teacher i up to t; if teacher i is part of the non-singleton team at some date, then there will be less information about his ability. Therefore, for characterizing the second-best, there is no loss of generality focusing on Ω_t^i.[4] Let $w_t^i(\Omega_t^i)$ be the wage paid to teacher i in period t. Then the constrained maximization problem to determine the second-best allocation

[4] Throughout we assume that contracts on current or future performance are not possible, so Ω_t^i contains all the contractible information.

can be written as:

$$\max_{\left\{w^i_{t+\tau}(\Omega^i_{t+\tau})\right\}_{\tau=0,1,..}} U_t^W \text{ subject to}$$

$$\{g_{t+\tau}\ ,\ b_{t+\tau}\}_{\tau=0,1,..} \in \arg \max_{\left\{g'_{t+\tau}\ ,\ b'_{t+\tau}\right\}_{\tau=0,1,..}} E_t\left[\sum_{\tau=0}^{\infty}\delta^{\tau}\left(w_{t+\tau}(\Omega^i_{t+\tau}) - g'_{t+\tau} - b'_{t+\tau}\right)\right]. \quad (9)$$

We leave the details of the analysis of the maximization problem in (9) to the Appendix. Here, note that any effort combination in the constraint set (i.e., any effort combination that is incentive compatible for the teacher) must satisfy:

$$\gamma f'\left(b_{t+\tau}\right) = f'(g_{t+\tau})$$

(see the Appendix). 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.

Next consider a wage schedule of the form $w_t = \alpha m_t + \kappa$, which links teacher compensation to their contemporaneous perceived ability (see the Appendix to see why focusing on such linear contracts is without loss of any generality). This formulation of incentives is similar to the seminal career concerns paper by Holmstrom [1999]. The extra effort put in by the teacher in period t increases her test score in period t. There are no immediate rewards for this increase as the teacher has already been paid her wage. However, an increase in the test score at t raises her perceived ability in period $t+1$ due to the updating rule (6). Moreover, because of the recursive nature of (6), the increase in perceived ability in $t+1$ has a (progressively dampened) ripple effect on all future expected abilities. Hence the present discounted value of the marginal benefit of higher test scores in period t can be summarized as: $\alpha\delta\beta[1+\delta(1-\beta)+\delta^2(1-\beta)^2+....] = \frac{\alpha\delta\beta}{1-\delta(1-\beta)}$. Notice that the marginal benefit of a higher test score is increasing in β, which is the coefficient on an individual teacher's test score in the ability updating rule: greater β implies that teacher effort will have a larger influence on future perceptions of his ability, and thus greater future rewards (which is the reason why the discount factor δ also matters). We thus define β as the "career concerns coefficient". The marginal benefit is also increasing in α, 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 privately optimal levels of good and bad effort, therefore, are:

$$f'(g_{t+\tau}) = \gamma f'(b_{t+\tau}) = \frac{1-\delta(1-\beta)}{\alpha\delta\beta}$$

for all $\tau \geq 0$. This implies that a greater α, i.e., 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 both on the career concerns coefficient β and the discount factor δ.

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

Proposition 1 The second-best solution is given by $g_{t+k} = g^{SB}$, and $b_{t+k} = b^{SB}$ for all k, with $g^{SB} < g^{FB}$. When each child is taught by a single teacher, the optimal wage schedule is given by $w_t^i = \alpha^{SB} m_t^i + \kappa$ for

$$\alpha^{SB} = \frac{1-\delta(1-\beta)}{\delta\beta f'(g^{SB})}, \tag{10}$$

and for any nonnegative κ. Both g^{SB} and α^{SB} are monotonically decreasing in γ , and we have $\alpha^{SB} < 1$, for $\gamma > \underline{\gamma}$.

Proposition 1 highlights the trade-off that the 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 γ captures the cost of higher incentives in the form of bad effort. Hence, an increase in γ increases the scope for bad effort and reduces the second-best level of good effort, g^{SB}, and consequently, the optimal level of incentives for the teacher, α^{SB}.

Equivalently, Proposition 1 can be understood in terms of two different types of negative externalities created by bad effort. The first is driven from the fact that since the effort levels are directly unobservable, the market's expectation of any individual teacher's effort level (which they back out in equilibrium) is based on the expectation of the market as a whole: when a teacher (or a positive mass of teachers) is expected to exert bad effort, a *given* test score translates into a lower perception of ability for other teachers. More explicitly, we have $m_{t+1}^i = m_t^i + \beta(z_t^i - m_t^i) - \overline{\beta}(\overline{z}_t^{-i} - \overline{m}_t^{-i})$ and

$z_t^i = s_t^i - f(g_t) - \gamma f(b_t)$ where b_t is the level of bad effort that teachers are expected to exert. Greater b_t reduces z_t^i, thus the perceived ability of other teachers. This negative externality is at the root of the inefficiency of various organizations. The second externality is driven by the presence of relative performance evaluation to back out η_t. Such evaluation creates a more direct negative externality from the actual level of bad effort by a teacher (as opposed to the expectation of bad effort): as a teacher exerts more bad effort, she increases parents' posterior about the common shock, η_t, and reduces their posterior about other teachers' abilities. When $n \to \infty$, this second externality is driven down to zero (since the common shock is completely revealed).[5]

Expression (10) ensures that the market-reward coefficient, α, is at the right level to ensure an effort level of g^{SB} given the career concerns coefficient implied by Bayesian updating, (7).[6] The following corollary emphasizes that second-best effort can also be achieved by manipulating β (if this were possible) for a fixed level of α:

Corollary The second-best equilibrium can alternatively be described by fixing α, and setting the career concerns coefficient on an individual teacher's test score equal to

$$\beta_{SB} = \frac{1-\delta}{\delta(\alpha f'(g^{SB}) - 1)}. \tag{11}$$

This discussion highlights two different channels via which the second-best allocation can be obtained. The first is by manipulating α, i.e., how the market rewards "success," and the second is by manipulating β, i.e., 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.

[5] This argument also shows that when $n < \infty$ and n declines, the externality becomes stronger and the gap between the second best and the first best widens.

[6] This discussion raises the possibility of beneficial government regulation directly manipulating α, for example, by tax policy. We do not consider this possibility since differential taxation of income from different occupations is rare in practice, and potentially very costly for a variety of reasons, including the distortions of such tax policies on the allocation of resources and talent across occupations.

Consider first the textbook model of perfectly competitive markets. Every teacher works independently, teaches a single child, and sells her teaching services in the market in every period. There is perfect competition among parents for education, and as a result each teacher gets paid her full expected output, given by the human capital equation (3). Wage w_t^i is given by

$$w_t^i = m_t^i + E_t[f(g_t^i)]. \tag{12}$$

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

Proposition 2 The market equilibrium is characterized by good effort level g^M, where

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

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

The proof follows from Proposition 1. The result that $g^M < g^{SB}$ 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 non-market institutions (e.g. tournaments) that may strengthen incentives even further, though we do not focus on those here.[7] 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, β, and via this, on σ_θ^2 and σ_ε^2. When σ_θ^2 is small relative to σ_ε^2, β 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

[7] 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].

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, i.e., when $\gamma > \underline{\gamma}$.

4 Incentives in firms

We now consider how firms can overcome these problems by creating teams of teachers to weaken the signaling ability of individual teachers.[8] We model the firm as a partnership of K teachers working together, engaged in joint teaching as captured in equation (2). Crucially, as equation (4) above shows, parents only observe the aggregate or average test score of all the teachers (or pupils) in the firm. Therefore, an important function of firms in one economy is to shut down the individual signals (test scores) of teachers.

More specifically, consider an allocation where there are J firms in the economy with the jth firm made up of K^j teachers, so that $\sum_{j=1}^{J} K^j = n$.[9] As before, along the equilibrium path parents can back out the signal $\overline{z}_t^j = \overline{a}_t^j + \overline{\theta}_t^j + \eta_t$ from $\overline{s}_t^j$ (again, with $J \to \infty$, η_t is backed out perfectly). Let

$$\overline{m}_t^j = \frac{1}{K^j} \sum_{i \in \mathbb{K}^j} m_t^{ji} \tag{13}$$

be the expected ability of the teachers in firm j at time t. Then parents update their belief about teacher i's (working in firm j) ability according to an updating formula similar to (6):

$$m_{t+1}^{ji} = m_t^{ji} + \beta_F(\overline{z}_t^j - \overline{m}_t^j) - \overline{\beta_F}(\overline{z}_t^{-j} - \overline{m}_t^{-j}). \tag{14}$$

Although parents can only observe the average test score of all the teachers in the firm, it is in theory possible for those inside the firm to have more information about each individual teacher's performance. We assume that, in addition to the average test score of all the teachers in the firm, insiders also observe the following signal of each teacher's

[8] We limit the analysis to the implicit incentives provided by firms. In addition, firms could improve the allocation of resources by providing explicit incentives, i.e., 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.

[9] There is a slight abuse of notation here, since before we were using j to refer to pupils, and now j refers to firms. This should cause no confusion.

performance (test score):

$$s_t^{ji} = a_t^{ji} + f(g_t^{ji}) + \gamma f(b_t^{ji}) + \theta_t^{ji} + \widetilde{\theta}_t^{ji} + \eta_t, \tag{15}$$

where $\widetilde{\theta}_t^{ji}$ is a normal error term, distributed as $N(0, \sigma^2_{\widetilde{\theta}})$. When $\sigma^2_{\widetilde{\theta}} \to \infty$, so that $\widetilde{\theta}_t^{ji}$ has a very large variance, we obtain the case where insiders observe exactly the same information as outsiders — i.e. there is no "asymmetry of information"; m_t^{ji}'s in (14) give the common in posteriors. We will start with this case of no asymmetry of information between insiders and outsiders, and then analyze the case where insiders have better information than the market.

Bertrand competition between parents again ensures that a group of teachers will be paid their expected contribution to human capital. Thus the *average* earnings of a teacher in firm j is:

$$\overline{w}_t^j = \overline{m}_t^j + \overline{f(g_t)}^j, \tag{16}$$

where $\overline{m}_t^j$ is given by (13), and the total revenue of the firm is $K^j \overline{w}_t^j$.

We also need to know how each individual teacher is rewarded (i.e., how the total revenue $K^j \overline{w}_t^j$ is divided between the teachers). We assume that each teacher is makes a take-it-or-leave-it offer to the partnership, so his wage will reflect his true contribution to the firm's revenue, thus:

$$w_t^{ji} = m_t^{ji} + \overline{f(g_t)}^j \tag{17}$$

where m_t^{ji} is the expectation of the ability of teacher i in firm j given the insiders' information set, with evolution given by (14). This wage rule parallels the market wage rule, (12), thus making it clear that the advantage of firms does not come from manipulating the wage rule, but from obscuring information.

We also assume that the set of teachers in firm j, $\mathbb{K}^j$, is chosen at time t and is not changed thereafter. In other words, teachers do not switch "teams" after initial assignment.[10] Finally, we assume that the firm (the partnership) maximizes revenues

[10] It can be shown that, as long as we are in the case with $\gamma > \underline{\gamma}$, "no-switching" is a long-run equilibrium. To see this briefly, consider a symmetric long-run equilibrium. According to (17), every teacher is paid her expected output equal to $w_t^{ji} = m_t^{ji} + \overline{f(g_t)}^j$. Hence in the long-run equilibrium, where g is constant, the future expected utility of a teacher if she stays in the firm is given by: $U_t^{ji} = E_t^j \left[\sum_{\tau=0}^{\infty} \delta^\tau \left(m_{t+\tau}^{ji} + f(g) - g - b \right) \right]$. Moreover, since ability is a random walk, we have $E_t^j[m_{t+\tau}^{ji}] = m_t^{ji}$, and hence $U_t^{ji} = \left(\frac{m_{t+\tau}^{ji} + f(g) - g - b}{1-\delta} \right)$. We next discuss deviations to switch to another team and to

minus the effort costs of the teachers. Then, the maximization problem can be written as:

$$\max_{\mathbb{K}^j} \max_{\left\{g^{ji}_{t+\tau}, b^{ji}_{t+\tau}\right\}_{i\in\mathbb{K}^j}} E^j_t \left[\sum_{\tau=0}^{\infty} \sum_{i\in\mathbb{K}^j} \delta^\tau \left(\overline{w}^j_{t+\tau}(K^j) - g^{ji}_{t+\tau}(K^j) - b^{ji}_{t+\tau}(K^j) - m^{ji}_t \right) \right], \tag{18}$$

where $\overline{w}^j_{t+\tau}$ is given by (16), K^j denotes the total number of teachers in that firm, and E^j_t is the expectations given the information set of the insiders in firm j at time t.[11] $\overline{w}^j_t$, $g_{t+\tau}$,and $b_{t+\tau}$ are written as functions of K^j to emphasize that the size of the firm will influence incentives and payments. In addition, the maximization is over the choice of a set $\mathbb{K}^j$, and finally, the term $\sum_{i\in\mathbb{K}^j} m^{ji}_t$ is the sum of the outside options of the teachers and also acts as a convenient normalization. This normalization makes it clear that the maximization is identical to the simpler maximization problem over the size of firms, K^j. Or more precisely,

$$\max_{K^j} \max_{\left\{g^{ji}_{t+\tau}, b^{ji}_{t+\tau}\right\}_{i\in\mathbb{K}^j}} E^j_t \left[\sum_{\tau=0}^{\infty} \delta^\tau \left(\overline{w}^j_t(K^j) - g^j_{t+\tau}(K^j) - b^j_{t+\tau}(K^j) \right) \right]. \tag{19}$$

---

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 $E^j_t[m^{ji}_{t+\tau}] = m^{ji}_t$, 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 old co-workers. 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 can also verify that when no other teacher is switching in the economy, a deviation to switching to another team 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, $E_t\left(m^{ji}_t\right)$, which can be more than $E^j_t\left(m^{ji}_t\right)$, the expectation of ability given the firm j information set, introducing an adverse selection problem. While a full analysis of adverse selection in this context is beyond the scope of our paper, we can see that since there is no the switching, a reasonable set of off-the-equilibrium path beliefs would be that switchers have arbitrarily low $E^j_t\left(m^{ji}_t\right)$ (i.e., they are selected from those with the most to gain from a deviation), making no-switching an equilibrium.

[11] Modeling the firm as a partnership rather than a profit-maximizing entity employing teachers is useful in simplifying the problem to a single maximization problem as in (18). With a profit-maximizing firm, the equilibrium would be characterized by a constrained maximization problem, whereby the firm takes the optimal effort choices of the teachers given their wage functions as given in its maximization.

One disadvantage of the partnership formulation, however, is that $\frac{1}{K^j}\Sigma_{i\in\mathbb{K}^j} w^{ji}_t \neq \bar{w}^j_t$, though naturally $E^j_t\left(\frac{1}{K^j}\Sigma_{i\in\mathbb{K}^j} w^{ji}_t\right) = E^j_t\left(\bar{w}^j_t\right)$, so the actual earnings of teachers will also include a mean-zero random element. Although all agents are risk-neutral, we can also think of this random element being perfectly insured across firms in the economy.

4.1 Symmetric information—$\sigma_{\tilde{\theta}}^2 \to \infty$

In this case, the firm can only make payments to teachers conditional on the past history of $\overline{s}_t^j$, the average signal from all the teachers (as well as the initial prior about the individual teacher). To see the benefits of large firms in the simplest possible way, let us return to the updating equation, (14), which is similar to the updating equation in the market case, (6). 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 $\frac{\beta_F}{K}$. 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 $\frac{1}{K}$ of the value created for the team Moreover, as the proof to Proposition 3 in the Appendix will show, $\beta_F = \beta$, and $\overline{\beta_F} = \overline{\beta}$. Since $\frac{\beta}{K}$ 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 $\frac{\beta}{K^*} = \beta_{SB}$ where β_{SB} is the career concerns coefficient that would ensure the second-best with $\alpha = 1$, as defined by equation (11). Moreover, given that the maximization problem in (19) is identical to the social surplus maximization problem (8), the firm will select $K = K^*$, and the second-best outcome is attained. This discussion establishes the following result:

Proposition 3 Suppose that $\sigma_{\tilde{\theta}}^2 \to \infty$. Then for a firm of size K, the good effort level chosen by a teacher, g, is given by $g^F(K)$, where g^F is monotonically decreasing in K with $g^F(1) = g^M$ and $g^F(K) \to 0$ as $K \to \infty$.

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

The proof is given in the Appendix. Here we outline the intuition behind the result. As in the market equilibrium, a teacher is still paid her expected output. However, the marginal effect of test score at time t on future expected ability is lower in firms than in markets. In other words, firms lower the career concerns coefficient from β to $\frac{\beta}{K}$, thus weakening individual incentives.

The reduction of career concerns effects under firms can thus completely redress the "over-incentivization" problem. With firms of appropriate size, the second-best allocation of Section 2.3 is achieved. When firms compete to maximize their value, all firms will endogenously expand to the optimal size K^*, since by definition K^* gives the maximum of (19). As a result, firm size here is a method of dulling incentives.

4.2 Asymmetric information and commitment—$\sigma^2_{\tilde{\theta}} < \infty$

The idea that the outside market only observes the average and not individual test scores of teachers is meant to capture the idea that the organizational structure of firms can be used to mask or suppress information. The question still remains, however, as to what extent the firm as a whole (or the principal/the owner), has access to information regarding an individual teacher's test score. Proposition 3 above implicitly assumed that nobody in the firm is able to observe individual teachers' test scores either. Alternatively, the allocation in Proposition 3 can be achieved if firms can announce a wage contract of the form (17), i.e., one that does not make any use of non-publicly available information, for all of their employees, and make a strict commitment to (not renegotiating) this wage contract.

We now relax the assumption of symmetric information by assuming that $\sigma^2_{\tilde{\theta}} < \infty$. This implies that insiders now observe a noisy signal of individual teacher performance as well as the public signal coming from average firm performance. In addition, we also assume that firms cannot commit to not modifying the rewards of their employees if this is in their interests. This is plausible given the various ways in which firms can enter into side deals with their employees. Notice that without the asymmetry of information, firms had no ability to manipulate incentives by modifying employee rewards, so there was no need for firms to commit to wage contracts, hence no commitment problem. The commitment problem is introduced by the asymmetry of information.

With the asymmetry of information of this sort, and the resulting commitment problem, firms become less attractive because, given the *ex post* manipulation of incentives inside the team, there will be limits to how much they can reduce the power of the incentives. More specifically (see Appendix for the proof):

Proposition 4 Suppose $\sigma^2_{\tilde{\theta}} < \infty$. There exists $\overline{\sigma^2_{\tilde{\theta}}}$, such that when $\sigma^2_{\tilde{\theta}} > \overline{\sigma^2_{\tilde{\theta}}}$ and $\gamma > \underline{\gamma}$, there is a unique equilibrium in which firms have size equal to $K^{**}\left(\sigma^2_{\tilde{\theta}}\right) > 1$, where $K^{**}\left(\sigma^2_{\tilde{\theta}}\right)$ induces the second-best level of effort g^{SB}, and is decreasing in $\sigma^2_{\tilde{\theta}}$. When $\sigma^2_{\tilde{\theta}} \leq \overline{\sigma^2_{\tilde{\theta}}}$, the second-best outcome cannot be achieved. When $\sigma^2_{\tilde{\theta}} = 0$, the firm equilibrium leads to the market outcome, i.e., the good effort level g^M.

As internal signals become more precise, a profit-maximizing firm will always use that extra information to encourage teachers to exert further bad effort, thus improving outside perception of the average ability of its employees, and via this channel, its future revenues. This implies that 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, i.e., typically, $\sigma^2_{\tilde{\theta}} > 0$, 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 $K^{**}\left(\sigma^2_{\tilde{\theta}}\right)$, 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 $\sigma^2_{\tilde{\theta}} \leq \overline{\sigma^2_{\tilde{\theta}}}$ for some critical threshold $\overline{\sigma^2_{\tilde{\theta}}}$, then there is sufficiently good internal information about teacher performance that even a very large firm would not be able to dull incentives sufficiently. Therefore, in this case the asymmetry of information, and the resulting commitment problem, breaks the "firm equilibrium" of Proposition 3, which achieved the second-best.

The intuition for why the asymmetry of information and the associated commitment problem 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 the profits. The problem is whether these high-powered incentives of the firm will trickle down to employees. With symmetric information, the firm has no way of increasing the incentives of its employees, thus there is no trickle-down of its high-powered incentives. The asymmetry of information introduces the possibility that the firm can manipulate its employees' incentives, and this, combined with the inability to commit to observable contracts, makes its high-powered incentives trickle down to the teachers. As a result, the benefits of firms in terms of dulling incentives are reduced or disappear.

5 Incentives in Governments

The analysis above highlighted the potential role of firms in improving efficiency through their ability to suppress information. However, it also pointed out the limitations of firms to credibly commit to such a course of action in the presence of the informational asymmetries between the firm and the outside world. Let us now imagine a world with $\gamma > \underline{\gamma}$, so that markets provide too high-powered incentives, and with $\sigma_{\tilde{\theta}}^2$ small so that firms cannot commit to dulling individual incentives because of severe asymmetries of information between insiders and outsiders. In particular, let us assume $\sigma_{\tilde{\theta}}^2 = 0$, which implies that firms will be unable to solve the over-incentivization problem, and g^M (the market equilibrium) is also the firm equilibrium.

Governments are widely believed to offer particularly 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]). Another possibility is that the public fears that politicians would steal from the public by appointing or promoting cronies or by demanding bribes from civil servants in exchange for favorable treatment, and hence requires politicians to obey constitutional or civil service rules that offer little scope for offering strong incentives to civil servants. In our model, we can think of these concerns imposing a wage structure on government organizations of the form $w_t^i = \alpha^G m_t^i + \kappa$. If α^G were close to α^{SB}, i.e., 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 these possibilities, our model suggests several other reasons why governments may be particularly able to commit to low-powered incentives. We now discuss these issues using a highly stylized model of government organization whereby the government (a politician) decides the size of schools and individual teacher rewards. This politician is potentially self-interested. Moreover, similar to individual teachers in our analysis so far, she would like to convince the market (i.e. the voters) that she has high ability, for example, to increase his re-election probability.

Denote the politician's true ability and market's perception of it at time t by a_t^p and m_t^p respectively. Further suppose that the politician has an objective function

$$U_t^{pol} = E_t \left[\sum_{\tau=0}^{\infty} \delta^\tau \left(m_{t+\tau}^p - C_{t+\tau} \right) \right], \tag{20}$$

where $C_{t+\tau}$ is the cost per student of the schooling system, or $C_{t+\tau} = w_{t+\tau}^{ij}$. 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. Suppose also that the ability of the politician evolves according to

$$a_{t+1}^p = a_t^p + \varepsilon_t^p,$$

where ε_t^p is i.i.d. with $\varepsilon \backsim N(0, \sigma_p^2)$ just as in equation (1).

Given this setup, we discuss three different reasons for why governments may be better able to commit to low-powered incentives than markets and firms. First, if there are "no politician effects", the ability or actions of the politician do not matter directly for student performance (other than through the incentives provided to teachers). Hence there is no commitment problem and politicians can enforce the second best level of effort. Second, if there are "politician effects", then even though the politician has an incentive to inflate test scores, those incentives may be lower than those in firms in the presence of common shocks because of the absence of yardstick competition in governments. Third, even in the absence of common shocks, governments may still have lower career concerns because of other sources of re-election uncertainty that politicians face.

5.1 Government operation with "no politician effects"

We start with the case where the politician's ability or action does not matter for student performance—except through the politician's manipulation of teachers' incentives. This immediately implies that no action that the politician can take will send a positive signal about her ability. This removes the commitment problem faced by firms. Therefore, the politician can choose, and commit to, the allocation of Proposition 3. In particular, she can set up $J = n/K^*$ schools of size K^* each, and promise a payment of $\overline{w}_t^j = \overline{m}_t^j + \overline{f(g_t)}^j$ to each teacher, replicating the allocation of Proposition 3, which, as shown earlier, coincides with the second-best. Alternatively, the politician can simply set the wage of each teacher equal to $w_t^i = \alpha^{SB} m_t^i + \kappa$.

This is true even when the politician has superior information about teacher test scores relative to the market, i.e., he observes a signal s_t^{jk} like in (15) with $\sigma_{\tilde{\theta}}^2 < \infty$. Although she can reward individual teachers on the basis of this additional signal, s_t^{jk}, and encourage them to exert more good and bad effort, the resulting increase in test scores will not lead to a better voter belief about her ability, since a higher average test score will not constitute a positive signal about ability, a_t^p. As a result, the politician has no reason to manipulate teacher incentives. This enables government organizations to commit to low-powered incentives.

5.2 Government operation with common shocks

The above analysis may be criticized because there is no room for the actions of the politician to influence outcomes other than through her effect on teacher effort. This means that the politician has no incentive to "inflate" student performance, and could easily provide, and commit to, the second-best incentives for teachers. 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 in school j is

$$h_t^{ji} = \overline{a}_t^j + \lambda a_t^p + \overline{f(g_t)}^j, \tag{21}$$

where a_t^p is the ability of the politician in charge of the schooling system, and $\overline{a}_t^j$ and $\overline{f(g_t)}^j$ are defined as above. This formulation implies that the politician's ability 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 his decisions. Consequently, the politician has an incentive to inflate test scores in order to improve others' perception of his own ability. It is straightforward to verify that the above analysis is unaffected by this modification and the term λa_t^p is now included in an augmented common shock, $\eta_t' = \lambda a_t^p + \eta_t$.

The point that we want to make 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, η_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, and update their beliefs regarding the ability of the politician according to the equation

$$m_{t+1}^p = m_t^p + \beta^p \left(z_t - m_t^p \right)$$

where

$$z_t = \frac{1}{n'} \sum_{i=1}^{n'} \overline{s}_t^i - \overline{A} - \overline{f(g_t)} - \gamma \overline{f(b_t)} = \lambda a_t^p + \eta_t$$

and

$$\beta^p = \frac{\lambda \left(1 + \sqrt{1 + 4 \left(\frac{\sigma_\eta^2}{\lambda^4 \sigma_p^2}\right)}\right)}{\lambda^2 \left(1 + \sqrt{1 + 4 \left(\frac{\sigma_\eta^2}{\lambda^4 \sigma_p^2}\right)}\right) + 2 \left(\frac{\sigma_\eta^2}{\lambda^4 \sigma_p^2}\right)}. \tag{22}$$

where n' is the number of firms in the economy, $\overline{A}$ is the average ability of teachers in the population, and $\overline{s}_t^i$ refers to the average test score of firm i at time t. These updating equations have an intuition similar to (6) and (7). 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 $\overline{A} + \overline{f(g_t)} + \gamma \overline{f(b_t)} + \lambda a_t^p + \eta_t$. The career concerns coefficient of the politician, β^p, is different from that of firms (or individual teachers), β, 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, η_t, not from the student performance shocks, the θ_t's. Here the absence of relative performance evaluation (yardstick competition) with government operation is important. The reason why σ_η^2 did not feature in the updating equations

(6) and (7) 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,[12] 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 his incentives, and therefore, indirectly those of the whole government organization. The greater σ_η^2, i.e., the more important the aggregate shock, the smaller β^p, and the weaker the incentives in governments. In the limit, as $\sigma_\eta^2 \to \infty$, 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 (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,

$$w_t^{ij} = \alpha_{t+\tau}^p m_{t+\tau}^{ij} + \kappa,$$

where κ is some constant. First, consider the case where the level of incentives provided to teachers $\alpha_{t+\tau}^p$ is observable. Then, even though the politician can manipulate teacher incentives, he 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 that with no politician effects, and the politician would simply choose $\alpha_{t+\tau}^p = \alpha^{SB}$ 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 $\alpha_{t+\tau}^p$'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 (20) by choosing the wage function of teachers. Since the government acts as a monopolist, it will only give each teacher their minimum reservation utility. Let $\underline{u}$ be the spot reservation utility of a teacher putting in zero effort. Then the government must pay each teacher a wage

[12] This argument needs to be qualified when local politicians run local school districts, for example, as in the U.S. 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.

equal to $(\underline{u} + g_t + b_t)$ each period where g_t and b_t are the effort levels that α^p induces. In other words, $w_t^{ij} = \alpha_t^p m_t^{ij} + \kappa = (\underline{u} + g_t^{ij} + b_t^{ij})$ in each period. Plugging the wage function into (20) and maximizing with respect to $\alpha_{t+\tau}^p$ gives us the following result:

Proposition 5 Suppose that $\sigma_{\tilde{\theta}}^2 = 0$ 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 by Proposition 2. The equilibrium level of effort under government operation, g^G, is then given by:

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

We have $g^G < g^M$ if and only if $\beta^p < \beta$. β^p is decreasing in σ_η^2 and increasing in λ.

The proposition establishes that government operation often provides weaker incentives than firms and markets, even when politicians have an interest in inflating test scores, and the manipulation of teacher incentives by the politician is not observed by voters. The reason is the presence of common shocks (hence the importance of σ_η^2), which were filtered out in markets, but not under government operation. The presence of common shocks increases the amount of noise in the performance of the politician, weakening his 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, i.e., when σ_η^2 is large, and when the contribution of the politician to aggregate test scores, λ, and the room for the politician to prove that he has high ability, σ_p^2, is 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 σ_η^2 is small, and/or when λ and σ_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.

5.3 Government operation under re-election 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. If so, perhaps similar outcomes could be achieved through a regulated private monopoly. However, in this subsection, we discuss another reason for lower-powered incentives with governments – re-election uncertainty.[13]

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

$$U_t^r = E_t \left[\sum_{\tau=0}^{\infty} (\delta^r)^\tau \left(m_{t+\tau}^r - C_{t+\tau} \right) \right], \tag{23}$$

where r denotes either to private monopoly or the politician, and the only difference from (20) is that the discount factor is denoted by δ^r. Our argument is that because politicians not only run schools but also control many other policies, and are subject to scandals, and external shocks, there will be reasons for re-election uncertainty uncorrelated with their performance in the task of running schools (i.e., correlated with $m_{t+\tau}^r$).[14] Assume simply that there is a probability $1 - \pi > 0$ that a politician will be disqualified and never be re-elected for these other reasons, in which case he receives 0 utility thereafter. Therefore, $\delta^{pol} = \delta\pi$. Suppose for comparison that $\delta^{mon} = \delta$, and also that $\beta^p = \beta$ so that without the issue of re-election uncertainty, governments would be identical to markets. We then have:

Proposition 6 Suppose that $\sigma_{\bar{\theta}}^2 = 0$ 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 by Proposition 2. Then in the absence of common shocks but under re-election uncertainty, the level of effort under government operation g^G is given by:

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

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

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

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

[13] It can be argued that limiting yardstick competition through government operation [CHECK] may be more feasible. 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.

[14] Although the CEO of a private firm can also be fired for events unrelated to his job-ability, such uncertainty is greater for politicians.

Higher re-election uncertainty (lower π) therefore lowers the career-concerns and effort level of the politician. If there is too much uncertainty in the election process, then g^G may drop substantially below g^{SB}, and case government operation will no longer be better than firms or markets. This suggests that while some natural level of uncertainty in the electoral process may be helpful, if the 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 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.

First, it is useful to note that while 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 non-interest, non-military federal budget is spent by the Education Department, Social Security, and Health and Human Services.[15] 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 would choose to be involved in education rather than in the operation of factories.[16]

The Case of 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 sub-

[15] See U.S. Office of Management and Budget, historical tables.

[16] Bowles and Gintis [1976], Lott [1999], Kremer and Sarychev [1997], and Pritchett [2002] suggest that governments may run schools in order to control what ideology is taught to students. 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.

sidizing education, it is unclear why governments operate schools, rather than simply subsidizing them. Consistent with the model, incentives are weaker in government-operated schools, and there is evidence that high-powered incentives in teaching can create major distortions. Perhaps the cleanest such evidence is provided by 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. While 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 U. S. 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, Hollenback and Stone [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 which 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. 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 U.K. 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.

The Case of 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 comes 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 of 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 U.S.$10 a year per active affiliate to administer or 0.32 percent of annual covered earnings. In Chile, on the other hand, administrative costs average U.S.$51.6 a year or 1.70 percent of annual covered earnings. There is also evidence in Chevalier and Ellison [1999] that U.S. 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.[17]

[17] The long time periods involved in pensions and the presence of many unsophisticated investors make pensions more prone to signaling and quality-boosting advertising and increases the potential costs of high-powered incentives for pensions relative to many other types of financial intermediation.

The recent paper 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 non-informative" ads (another example of "bad" effort). Furthermore, the paper shows that such non-informative ads actually lead investors "astray" by exploiting their behavioral biases. Thus even in the pension fund context, the distortionary cost of higher incentives that this paper put forth can be quite large and important.

Other cases

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 make people feel better in the short-run but do not improve their long-run health. For example, U.S. hospitals provide more non-medical amenities than British hospitals, which face less competition. While it is certainly possible that British hospitals may be providing sub-optimal non-medical amenities, the evidence is also consistent with the notion that in the more market-based U.S. system, hospitals are trying to signal quality by providing easily observed non-medical 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 U.K. manages to achieve health outcomes nearly as good as the United States, while spending only 7.3% of GDP on health compared to the 13% the United States spends (OECD Health Data [2002]). Finally, law enforcement agents with too high-powered incentives may frame innocent people to appear more able to solve crimes, so regulation of incentives might again be necessary.

7 Concluding Remarks

This paper has presented a model in which both high and low-powered incentives have costs. While 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

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.

extent of incentives. The natural career concerns in market environments may then lead 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, reducing 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, re-election uncertainty due to other reasons weakens politicians incentives. Weaker politician incentives in turn imply 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.[18]

At the other extreme, government operation may be appropriate for tasks where it is

[18] 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 non-profits, 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 non-profit firms).

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 paper.

There are of course limits to the theory presented in this paper, since in actual practice, many other factors are undoubtedly important, and the boundaries of markets, firms, and governments 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 paper 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.

8 Appendix

Proof of Lemma 1: Although in the text we focus on the case where $n \to \infty$, here we solve for the general case with n finite first. Parents with prior m_t^i about teacher i observe the vector Z_t. Let v_t be the variance of m_t^i. Since m_t^i and Z_t are distributed normally, we can use the normal updating formula to update the ability prior and compute m_{t+1}^i conditional on Z_t. This formula is given by:

$$m_{t+1}^i = m_t^i + \Sigma_{12}\Sigma_{22}^{-1}(Z_t - M_t),$$

where $M_t = [m_t^1 \ldots\ldots\ldots m_t^n]^T$, $\Sigma_{12} = [0\ 0\ ..v_t....\ 0]$, with the convention that v_t corresponds to the ith component of the vector, and

$$\Sigma_{22} = \begin{bmatrix} (v_t + \sigma_\theta^2 + \sigma_\eta^2) & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & \ldots\ldots\ldots\ldots\ldots\ldots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & (v_t + \sigma_\theta^2 + \sigma_\eta^2) \end{bmatrix}.$$

Σ_{22} is an $(n\text{x}n)$ matrix with $(v_t + \sigma_\theta^2 + \sigma_\eta^2)$ as the diagonal term, and σ_η^2 as all the non-diagonal terms. Σ_{22}^{-1} can be written as:

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

where

$$\begin{aligned} b &= (n-1)\sigma_\eta^2 - \frac{(v_t + \sigma_\theta^2 + \sigma_\eta^2)^2}{\sigma_\eta^2} - (n-2)(v_t + \sigma_\theta^2 + \sigma_\eta^2), \text{ and} \\ a &= -\left[\frac{(v_t + \sigma_\theta^2 + \sigma_\eta^2)}{\sigma_\eta^2} + (n-2)\right]. \end{aligned}$$

Plugging in the value of Σ_{12} and Σ_{22}^{-1}, we obtain:

$$m_{t+1}^i = m_t^i + \beta(z_t^i - m_t^i) - \overline{\beta}(\overline{z}_t^{-i} - \overline{m}_t^{-i}), \tag{A1}$$

where:

$$\begin{aligned} \overline{z}_t^{-i} &= \frac{1}{(n-1)}\sum_{j \neq i} z_t^j \\ \overline{m}_t^{-i} &= \frac{1}{(n-1)}\sum_{j \neq i} m_t^j \end{aligned}$$

$$\beta = \frac{v_t a}{b} \text{ and } \overline{\beta} = -\frac{v_t(n-1)}{b}.$$

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

Next, we need to solve for v_t. Like m_t^i, v_t is also updated each period after the realization of Z_t. This updating formula is given by:

$$v_{t+1} = v_t - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \sigma_\varepsilon^2,$$

where Σ_{12} and Σ_{22} are defined above and $\Sigma_{21} = [\Sigma_{12}]^T$. Therefore, we have

$$v_{t+1} = v_t - \frac{a}{b}v_t^2 + \sigma_\varepsilon^2$$

Since we are interested in the long-run stationary equilibrium, we impose the condition $v_{t+1} = v_t = v$. This stationarity condition implies:

$$v^2 = \frac{b}{a}\sigma_\varepsilon^2$$

which can be expanded into:

$$v^3 + v^2[(n-1)\sigma_\eta^2 + (\sigma_\theta^2 - \sigma_\varepsilon^2)] - v[n\sigma_\eta^2\sigma_\varepsilon^2 + 2\sigma_\theta^2\sigma_\varepsilon^2] - [\sigma_\varepsilon^2(\sigma_\theta^2)^2 + n\sigma_\eta^2\sigma_\theta^2\sigma_\varepsilon^2] = 0 \quad \text{(A2)}$$

We now look at the case where $n \to \infty$. Then (A2) simplifies to:

$$v^2 - v\sigma_\varepsilon^2 - \sigma_\theta^2\sigma_\varepsilon^2 = 0,$$

which can be explicitly solved as:

$$v = \frac{\sigma_\varepsilon^2 + \sqrt{\sigma_\epsilon^4 + 4\sigma_\theta^2\sigma_\varepsilon^2}}{2}. \quad \text{(A3)}$$

Plugging (A3) back into (A1), and taking the limit and $n \to \infty$, we obtain the updating equation (A1), $(\overline{z}_t^{-i} - \overline{m}_t^{-i}) \to \eta$, and $\overline{\beta} \to \beta$, thus $\overline{\beta}(\overline{z}_t^{-i} - \overline{m}_t^{-i}) \to \beta\eta$, and we have

$$m_{t+1}^i = m_t^i + \beta\ (z_t^i - m_t^i - \eta)$$

where

$$\beta = \frac{v}{v + \sigma_\theta^2} = \frac{1 + \sqrt{1 + 4\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right)}}{1 + 2\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right) + \sqrt{1 + 4\left(\frac{\sigma_\theta^2}{\sigma_\varepsilon^2}\right)}}. \quad \text{(A4)}$$

It can be easily checked from (A4) that $0 < \beta < 1$, $\frac{\partial \beta}{\partial \sigma_{\varepsilon}^2} > 0$, and $\frac{\partial \beta}{\partial \sigma_{\theta}^2} < 0$.■

Proof of Proposition 1: The second-best is given by the solution to (9). Notice that (9) is a point-wise maximization problem over time. The constraint can then be written separately for each τ as:

$$f'(g_{t+\tau}) \sum_{\tau'=0}^{\infty} \delta^{\tau'} \frac{\partial w_{t+\tau+\tau'}}{\partial s_{t+\tau}} = 1, \text{ and } \gamma f'(b_{t+\tau}) \sum_{\tau'=0}^{\infty} \delta^{\tau'} \frac{\partial w_{t+\tau+\tau'}}{\partial s_{t+\tau}} = 1 \tag{A5}$$

The above conditions can be combined to give $f'(g_{t+\tau}) = \gamma f'(b_{t+\tau})$ for all τ which implies that:

$$b_{t+\tau} = f'^{-1}\left(\frac{f'(g_{t+\tau})}{\gamma}\right). \tag{A6}$$

The inverse of $f'(x)$ exists due to the concavity of $f(x)$. Equation (A6) defines the *feasible* pairs of $(g_{t+\tau},\ b_{t+\tau})$ even when the wage function is not differentiable. Restriction to differentiable wage functions therefore does not change our second-best solution. Given (A6), we can simplify our maximization problem (9) into the unconstrained problem:

$$\max_{g_{t+\tau}} \sum_{\tau=0}^{\infty} \delta^{\tau} \left(\overline{A} + f(g_{t+\tau}) - g_{t+\tau} - f'^{-1}\left(\frac{f'(g_{t+\tau})}{\gamma}\right)\right). \tag{A7}$$

The above is a well-defined maximization problem with a unique global maximum. Moreover, because of the additive nature of (A7), at the optimum, $g_t^{SB} = g_{t'}^{SB} = g^{SB}$ for all t, t'. Differentiating (A7) with respect to g_t, we obtain:

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

Because of the concavity of $f(x)$, the expression on the RHS is greater than 1, implying $g^{SB} < g^{FB}$.

Given g^{SB}, it is easy to solve for the optimal wage structure. In fact there is a continuum of optimal wage structures, as long as they satisfy the condition (from (A5)), $f'(g^{SB}) \sum_{l=1}^{\infty} (\delta^l \frac{\partial w_{t+k+l}}{\partial s_{t+k}}) = 1$. This condition can be satisfied by the wage schedule $w_t^i = \alpha^{SB} m_t^i + \kappa$. To see this, note that from (6), we can write

$$m_{t+l} = (1-\beta)^l m_t + \beta(1-\beta)^{l-1} s_t + \beta(1-\beta)^{l-2} s_{t+1} + ... + \beta s_{t+l-1} + \text{constant}.$$

We can then write $\frac{\partial w_{t+k+l}}{\partial s_{t+k}} = \alpha\beta(1-\beta)^{l-1}$, which implies that at the second-best

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

Since g^{SB} is decreasing in γ , there exists a $\underline{\gamma}$ such that for $\gamma > \underline{\gamma}$, $\alpha^{SB} < 1$. ■

Proof of Proposition 3: With firms, parents observe J signals, represented by:

$$\overline{Z}_t = [\overline{z}_t^1\ \overline{z}_t^2 ....\overline{z}_t^J]^T,$$

where

$$\overline{z}_t^j = \overline{a}_t^j + \overline{\theta}_t^j + \eta_t.$$

The are J firms in the economy with each firm j having a size K_j. Then using the normal updating formula, ability for teacher k in firm j is updated using:

$$m_{t+1}^{jk} = m_t^{jk} + \bar{\Sigma}_{12}\bar{\Sigma}_{22}^{-1}(\overline{Z}_t - \overline{M}_t),$$

where $\overline{M}_t = [\overline{m}_t^1........\ \overline{m}_t^J]^T$, $\bar{\Sigma}_{12} = [0\ 0\ ..\frac{v_{jt}^F}{K_j}....\ 0]$, with $\frac{v_{jt}^F}{K_j}$ corresponding to the jth component of the vector, and

$$\bar{\Sigma}_{22} = \begin{bmatrix} (\frac{v_{1t}^F+\sigma_\theta^2}{K_1} + \sigma_\eta^2) & \sigma_\eta^2 & \sigma_\eta^2 \\ \sigma_\eta^2 & .... & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 & (\frac{v_{Jt}^F+\sigma_\theta^2}{K_J} + \sigma_\eta^2) \end{bmatrix}.$$

$\bar{\Sigma}_{22}$ is a $(J\text{x}J)$ matrix with $\left(\frac{v_{jt}^F+\sigma_\theta^2}{K_j} + \sigma_\eta^2\right)$ as the diagonal term, and σ_η^2 as all the non-diagonal terms. Now take the limit $n \to \infty$, which, since $K_j < \infty$ for all j, implies $J \to \infty$. Then $\overline{\Sigma}_{22}^{-1}$ can be written as:

$$\overline{\Sigma}_{22}^{-1} = \begin{bmatrix} \frac{K_1}{v_{1t}^F+\sigma_\theta^2+K_1\sigma_\eta^2} & 0 & 0 \\ 0 & \cdots\frac{K_j}{v_{jt}^F+\sigma_\theta^2+K_j\sigma_\eta^2}\cdots & 0 \\ 0 & 0 & \frac{K_J}{v_{Jt}^F+\sigma_\theta^2+K_J\sigma_\eta^2} \end{bmatrix}.$$

Plugging in the value of $\bar{\Sigma}_{12}$and $\overline{\Sigma}_{22}^{-1}$, we get:

$$m_{t+1}^{jk} = m_t^{jk} + \beta_F(\overline{z}_t^j - \overline{m}_t^j) - \overline{\beta_F}(\overline{z}_t^{-j} - \overline{m}_t^{-j})$$

where:

$$\overline{z}_t^{-j} = \frac{1}{(J-1)}\sum_{l\neq j} z_t^l$$
$$\overline{m}_t^{-j} = \frac{1}{(J-1)}\sum_{l\neq j} m_t^l.$$
$$\beta_F = \frac{v^F}{v^F+\sigma_\theta^2}$$

The variance is then updated by:

$$v_{j,t+1}^F = v_{jt}^F - \bar{\Sigma}_{12}\bar{\Sigma}_{22}^{-1}\bar{\Sigma}_{21} + \overline{\sigma}_\epsilon^2$$

where $\overline{\sigma}_\epsilon^2 = \frac{\sigma_\varepsilon^2}{K_j}$. Assuming stationarity, i.e. $v_{j,t+1}^F = v_{jt}^F = v_j^F$,we get:

$$v_j^F \to v = \frac{\sigma_\varepsilon^2 + \sqrt{\sigma_\epsilon^4 + 4\sigma_\theta^2\sigma_\varepsilon^2}}{2} \equiv v^F.$$

Notice that the stationary firm variance is the *same* as the individual teacher stationary variance under the market equilibrium. Similarly, as $J \to \infty$, $\beta_F \to \frac{v^F}{v^F+\sigma_\theta^2} \equiv \beta$, $\overline{\beta}_F \to \beta$, and $(\overline{z}_t^{-j} - \overline{m}_t^{-j}) \to \eta_t$. In other words, as $J \to \infty$, 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* k in firm j is given by $\frac{\beta}{K_j}$, and is decreasing in K_j.

>From this, it is straightforward to see that $g^F(1) = g^M$ and $g^F(K_j) \to 0$ as $K_j \to \infty$. 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 partners try to maximize:

$$\max_{K_j} E_t\left[\sum_{\tau=0}^{\infty}\delta^\tau\left(\overline{m}_{t+\tau}^j + f(g_{t+\tau}(K_j)) - g_{t+\tau}(K_j) - b_{t+\tau}(K_j)\right)\right]. \quad \text{(A8)}$$

As we saw in Proposition 1, (A8) is maximized at $K_j = K^*$, such that $g(K^*) = g^{SB}$, providing the second-best solution. ■

Proof of Proposition 4: Let $n \to \infty$ so that with each firm of finite size, $J \to \infty$. 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:

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

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

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

$$m_{t+1}^{ji} = m_t^{ji} + \left[\begin{array}{cc} v_t^F & \frac{v_t^F}{K} \end{array} \right] \left[\begin{array}{cc} (v_t^F + \sigma_\theta^2 + \sigma_{\widetilde{\theta}}^2) & \frac{(v_t^F + \sigma_\theta^2)}{K} \\ \frac{(v_t^F + \sigma_\theta^2)}{K} & \frac{(v_t^F + \sigma_\theta^2)}{K} \end{array} \right]^{-1} \left[\begin{array}{c} (z_t^{ji} - m_t^{ji}) \\ (\overline{z}_t^j - \overline{m}_t^j) \end{array} \right],$$

which implies

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

where

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

defines the career concerns coefficient with asymmetric information, superscript $-i$ refers to the average excluding the ith teacher, and

$$\overline{\beta}^{asy} \equiv \frac{v_t^F \sigma_{\widetilde{\theta}}^2 K^2}{(v_t^F + \sigma_\theta^2)((K-1)(v_t^F + \sigma_\theta^2) + K\sigma_{\widetilde{\theta}}^2)(K-1)}.$$

Once again the stationary variance v^F will be given by $v_{t+1}^F = v_t^F = v^F$, which after applying the normal updating formula is given by the implicit equation,

$$\frac{(v^F)^2(K-1)(v^F + \sigma_\theta^2) + (v^F)^2 \sigma_{\widetilde{\theta}}^2}{(v^F + \sigma_\theta^2)((K-1)(v^F + \sigma_\theta^2) + K\sigma_{\widetilde{\theta}}^2)} = \frac{\sigma_\varepsilon^2}{K}.$$

To emphasize dependence on firm size, let us write the career concerns coefficient above, β^{asy}, as $\beta^{asy}(K)$. Let K^{**} be the value of K that makes $\beta^{asy}(K) = \beta_{SB}$. Then we have that $\frac{\partial K^{**}}{\partial \sigma_{\widetilde{\theta}}^2} < 0$. 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 $\frac{\partial\beta^{asy}(K)}{\partial K} < 0$, and $\beta^{asy}(K=1) = \beta_M$, to establish $\beta^{asy}(K) = \beta_{SB}$, we simply need to show that $\lim_{K\to\infty} \beta^{asy}(K) < \beta_{SB}$. We have:

$$\lim_{K\to\infty} \beta^{asy}(K) = \frac{v^F}{v^F + \sigma_\theta^2 + \sigma_{\tilde{\theta}}^2},$$

and

$$\left(\frac{v^F}{v^F + \sigma_\theta^2 + \sigma_{\tilde{\theta}}^2}\right) < \beta_{SB} \Leftrightarrow \sigma_{\tilde{\theta}}^2 > \overline{\sigma_{\tilde{\theta}}^2} \equiv \left(\frac{v^F}{\beta_{SB}} - v^F - \sigma_\theta^2\right).$$

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

Proof of Proposition 5: The politician's problem can be defined as follows:

$$\max_{\alpha_{t+\tau}^p} U_t^{pol} = E_t\left[\sum_{\tau=0}^{\infty} \delta^\tau \left(m_{t+\tau}^p - w_{t+\tau}^{ij}\right)\right],$$

where

$$w_{t+\tau}^{ij} = \alpha_{t+\tau}^p m_{t+\tau}^{ij} + \kappa_{t+\tau} = (\underline{u} + g_{t+\tau}^{ij} + b_{t+\tau}^{ij})$$

$$m_{t+1}^p = m_t^p + \beta^p\left(z_t - m_t^p\right).$$

Keeping market expectations of $g_{t+\tau}^{ij}$ and $b_{t+\tau}^{ij}$ fixed, and maximizing over $\alpha_{t+\tau}^p$, we get the first order condition:

$$\left(f'(g_{t+\tau}^{ij})\frac{\partial g_{t+\tau}^{ij}}{\partial \alpha^p} + \gamma f'(b_{t+\tau}^{ij})\frac{\partial b_{t+\tau}^{ij}}{\partial \alpha^p}\right)\left(\frac{\delta\beta^p}{1-\delta(1-\beta^p)}\right) = \left(\frac{\partial g_{t+\tau}^{ij}}{\partial \alpha^p} + \frac{\partial b_{t+\tau}^{ij}}{\partial \alpha^p}\right).$$

We know that teachers' first-order conditions imply that $f'(g) = \gamma f'(b)$. This simplifies the above expression to:

$$f'(g^G)\left(\frac{\delta\beta^p}{1-\delta(1-\beta^p)}\right) = 1,$$

thus completing the proof.■

Proof of Proposition 6:

The proof is identical to that of Proposition 5, but with $\beta^p = \beta$, and δ replaced by $(\pi\delta)$.■

Bibliography

Acemoglu, Daron. [1998] "Credit Market Imperfections and Separation of Ownership from Control," *Journal of Economic Theory*, February, 78(2), 355-381.

Acemoglu, Daron and Thierry Verdier. [2000] "The Choice Between Market Failures and Corruption," *American Economic Review*, 90(1), 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," *Quarterly Journal of Economics*, 112(4), 1289-1332.

Barberis, Nicholas; Maxim Boycko; Andrei Shleifer; and Natalia Tsukanova. [1996] "How does Privatization Work? Evidence from the Russian Shops," *Journal of Political Economy*,104(4), 764-790.

Bates, Robert. [1981] *Markets and States in Tropical Africa*, Berkeley, CA: University of California Press.

Bernstein, Solange and Alejandro Micco. [2003] "Turnover and Regulation: The Chilean Pension Fund Industry"Central Bank of Chile, Working Paper No. 180.

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: Basic Books.

Chevalier, Judith and Glenn Ellison. [1999] "Career Concerns of Mutual Fund Managers," *Quarterly Journal of Economics*, May, 114(2), 389-431.

Coase, Robert. [1937] "The Nature of the Firm," *Economica*, 4(16), 386-405.

Cronqvist, Henrik. [2003]. "Advertising and Portfolio Choice", University of Chicago, GSB Mimeo.

Crozier, Maurice. [1967] *Bureaucratic Phenomenon*, University of Chicago Press.

Dewatripont, Mathias; Ian Jewitt; and Jean Tirole. [1999] "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, January, 66(1),183-198.

Diamond, Peter and Salvador Valdes-Prieto. [1994] "Social Security Reforms" in Bosworth, Barry P., Rudiger Dornbusch and Raul Laban, *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," *American Economic Review*, Papers and Proceedings, 87(2), May, 378-382.

Dixit, Avinash. [2002] "Incentives and organizations in the public sector: An interpretative review", *Journal of Human Resources* 37 (4), Fall, 696-727.

Eberts, Randall; Hollenbeck, Kevin; Stone, Joe. [2002] "Teacher Performance Incentives and Student Outcomes", *Journal of Human Resources*, Fall Vol. 37, Issue 4.

Figlio, David N. and Joshua Winicki. [2002] "Food for Thought: The Effects of School Accountability Plans on School Nutrition," National Bureau of Economics Research Working Paper: 9319.

Gibbons, Bob. [1998] "Incentives in Organizations," National Bureau of Economic Research Working Paper: 6695.

Gibbons, Bob and Murphy, Kevin. [1992] "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, June, 100(3), 468-505.

Glennerster, Howard. [2002] "United Kingdom Education 1997-2001," *Oxford Review of Economic Policy*, 18(2), 120-136.

Glewwe, Paul; Nauman Ilias; and Michael Kremer. [2003] "Teacher Incentives," National Bureau of Economic Research Working Paper: 9671.

Green, Jerry and Nancy Stokey. [1983] "A Comparison of Tournaments and Contests," *Journal of Political Economy*, 91(3), 349-64.

Grossman, S. and O. Hart. [1986] "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94(4), 691-719.

Harbison, Ralph W. and Eric A. Hanushek. [1992] *Educational Performance of the Poor: Lessons from Rural Northeast Brazil*, New York: Oxford University Press.

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.

Hart, O. and J. Moore. [1990] "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98(6), 1119-1158.

Hart, O.; A. Shleifer; and R. Vishny. [1997] "The Proper Scope of Government: Theory and an Application to Prisons," *Quarterly Journal of Economics*, November, 112(4), 1127-1161.

Holmstrom, B. [1979] "Moral Hazard and Observability" *Bell Journal of Economics*, 10, 74-91.

Holmstrom, B. [1999] "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, January, 66(1), 169-182.

Holmstrom, B., and Milgrom P. [1991] "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization*, 7 (Special Issue): 24-52.

Holmstrom, B. and P. Milgrom. [1991] "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, 7 (Special Issue), 24-52.

Holmstrom, B., and P. Milgrom. [1994] "The Firm as an Incentive Structure," *American Economic Review*, 84(4), 972-991.

Hsieh, Chang-Tsi and Miguel Urquiola. [2002] "When Schools Compete, How Do They Compete? An Assessment of Chile's Nationwide School Voucher Program," unpublished.

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." University of Chicago Press.

Kihlstrom, R. and M. Riordan. [1984] "Advertising As a Signal," *Journal of Political Economy*, 92(3), 427-450.

Kremer, Michael. [1997] "Why Are Worker Cooperatives So Rare?," National Bureau of Economic Research Working Paper: 6118.

Kremer, Michael and Andrei Sarychev. [1998] "Why Do Governments Operate Schools," unpublished.

Ladd, Helen and Edward Fiske. [2000] *When Schools Compete: A Cautionary Tale*, Washington, D.C.: The Brookings Institution.

Levin, J. and S. Tadelis. [2002] "A Theory of Partnerships," Stanford University Working Paper.

Lott, Jr., John. [1999] "Public Schooling, Indoctrination and Totalitarianism," *Journal of Political Economy*, 107(6), 127-157.

Meyer, M. and J. Vickers. [1997] "Performance Comparisons and Dynamic Incentives," *Journal of Political Economy*, June, 105(3), 547-581.

Milgrom P. and J. Roberts. [1986] "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94(4), 796-821.

Niskanen, William A. [1971] *Bureaucracy and Representative Government*, Chicago: Aldine/Atherton.

Prendergast, Canice. [2003] "The Limits of Bureaucratic Efficiency," forthcoming *Journal of Political Economy.*

Pritchett, Lant. [2002] "When Will They Ever Learn?: Why All Governments Produce Schooling," unpublished.

Rajan, R. and L. Zingales. [1998] "Power in a Theory of the Firm," *Quarterly Journal of Economics*, May, 113(2), 387-432.

Shleifer, A. [1985] "A Theory of Yardstick Competition," *Rand Journal of Economics*, 16(3), 319-327.

Shleifer, Andrei and Robert W. Vishny. [1994] "Politicians and Firms," *Quarterly Journal of Economics*, November, 109(4), 995-1025.

Stein, J. [1989] "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *Quarterly Journal of Economics*, November, 104(4), 655-669.

Tirole, Jean. [1986]. "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations". *Journal of Law Economics and Organization*, 2, 181-214.

Williamson, O. [1985] *The Economic Institutions of Capitalism*, New York: Free Press.

Wilson, James. [1989] *Bureaucracy: What Government Agencies Do and Why They Do It*, Basic Books Classics.