‘Public service motivation’ as an argument for government provision

Patrick Francois

Abstract

A public service motivation (PSM) inclines employees to provide effort out of concern for the impact of that effort on a valued social service. Though deemed to be important in the literature on public administration, this motivation has not been formally considered by economists. When a PSM exists, this paper establishes conditions under which government bureaucracy can better obtain PSM motivated effort from employees than a standard profit maximizing firm. The model also provides an efficiency rationale for low-powered incentives in both bureaucracies and other organizations producing social services. © 2000 Elsevier Science S.A. All rights reserved.

Keywords: Privatization; Bureaucracy; Public service motivation

JEL classification: H11; H42; J45; J32

1. Introduction

Public administration scholars use the term ‘Public Service Motivation’ (PSM) to refer to a “general, altruistic motivation to serve the interests of a community of people, a state, a nation or humankind, . . . ” Rainey and Steinbauer (1999, p. 20).
There is considerable survey-based evidence that such motivations exist. Some argue that these motivations should, and do, serve as the defining difference between public and private institutions (Denhardt, 1993; Fredrickson, 1997; Rainey and Steinbauer, 1999). They argue that effective and well-functioning public organizations are populated by individuals with a sense of PSM, that this sense actively motivates employees in their work, and, implicitly, that such considerations do not motivate employees of private firms. Though not formally shown, a theme in this literature is that government provision (in addition to purchase) of certain public services is efficient, since it allows an avenue for this service ethic to be expressed and lowers the price to taxpayers, see Holtham and Kay (1994). Under this view, the well-documented differences in incentive-based payment between public and private sector managers (Rainey, 1983) is thought to be compensated by a non-pecuniary benefit for the public employee, i.e. meeting their altruistic goals. Public sector employees, who are motivated by PSM, are thus seen as volunteering a portion of their services to the community for free, Gassler (1998). Such voluntary provision, it is conjectured, would not be forthcoming if working for a private, profit maximizing, firm:

“The logic of service and the logic of profit are largely at odds. Why volunteer your time to an organization whose entire purpose is making a buck?” Kuttner (1989, p. 24).

At first glance, most economists would think this informal argument wrong. Employees motivated by PSM, are concerned with outcomes. If their effort contributes to realizing the outcomes they care for, then their motivation to contribute part of that effort voluntarily should be independent of their employer’s identity.

This paper shows, in contrast, that the employer may matter, and that the informal arguments made in favour of government provision in the public administration literature do have economic merit. Consider the problem facing a social planner who wants to ensure provision of a valuable social service. The planner chooses between setting up a government department, which has a bureaucrat hired to oversee government employees, or writing a contract with a private firm to provide the service. The main distinction between a contract with a private firm and the incentive structure in a bureaucracy is that the firm’s owner is a residual claimant on net profit, whereas a bureaucrat in the public sector is not. Suppose all workers are endowed with PSM, and that this concern exists independently of who they work for. This is an outcome-oriented motivation so the workers do not care who provides the service, and do not get utility from the

---

very act of providing it, rather they only care about the level of service. For PSM to motivate a worker’s own effort, the worker must believe that, were she not to provide the effort, the level of service would fall. The important thing then is for the employer (either the government bureaucrat or residual claimant in the private firm) to credibly commit not to make up for any shirking by adjusting other inputs. When this commitment is credible, the worker will not shirk since she knows shirking affects output. The private firm cannot credibly commit because once shirking occurs, profits will be lost if it does not adapt and the contract with the planner is unfulfilled. The government bureaucrat, however, with no residual claim on profits, does not face a similar incentive. Under a range of conditions, which are established in the paper, the bureaucrat can commit not to adapt, so that government workers are motivated by PSM not to shirk even though workers in a private firm would not be.

Under these conditions, a government department provides the service at strictly lower cost than a profit-maximizing firm operating under contract with the planner. The analysis bears some similarity to a number of different theories that argue a case for public provision and it is worth considering these distinctions before proceeding.

Psychologists have noted that the perceived nature of a relationship affects individual motivations. Consequently, some argue that civic virtues can be crowded out when services provided as gifts are then provided under contract. In particular, Gregory (1999), Frey (1997a,b) and Nowland-Foreman (1998) argue that the form of contracting in the public sector affects individuals’ norms. When contracts treat agents as if they were self-serving, they then become so. Here, in contrast, the level of PSM in the population is assumed to be unchanged by the form of contracting and service provision. Preferences of all agents are identical whether working for a private firm or government, and all individuals are homogeneous. This also contrasts the present paper with work examining adverse selection problems that arise when individuals differ in their degree of altruistic motivation, see Weisbrod and Schlesinger (1986) in Rose-Ackerman (1986).2

2 Preston (1989) also develops a model analyzing the self-selection of those individuals with high social valuations into non-profit employment. The concerns that arise from heterogeneity in preferences will not be examined here. The robustness of results to assuming heterogeneity would depend on the information about heterogeneity in the population. If individuals differed in PSM, but this was publicly observable, then all results would simply go through as is, with the high-PSM individuals working in government. Alternatively, if PSM is privately known, but there is sorting through time based on observation, then the model here would correspond to an equilibrium of a more complicated model in which, in steady state, individual types are revealed. If there is no opportunity for observation or sorting, government would always be subject to opportunist behavior on the part of individuals pretending to have high PSM. In that case, there would be a tradeoff between the benefits to having employees with high PSM and the costs of being cheated by low-PSM employees. The results here would go through only if these costs were not too great. To reiterate, however, the present paper assumes that all individuals have equivalent PSM, and that this is commonly known.
An alternative, though related argument, for government provision, which draws on the multi-task principal agent literature developed by Holmstrom and Milgrom (1991) has been made by Hart et al. (1997). Where there exist non-contractible elements of service provision, they argue private firms have incentive to undertake non-verifiable cost-reducing actions which compromise the quality of provision. In contrast, a government, where savings are not pocketed by bureaucrats, has no incentive to compromise quality and lower costs in this way. A similar argument has been made by Glaeser and Shleifer (1998) to explain the existence of NPOs (Non-profit organisations) when output, or the service provided, is not fully contractible with the purchaser. The theory presented here, in contrast, assumes full and perfect contractibility over the service provided but depends on the existence of PSM. Thus, the private contractor has no recourse to profit by lowering the quality of service provision, so that the considerations in Hart et al. (1997) are not operating here.

Although the output of service is assumed to be fully contractible in the analysis here, a critical component of the theory is that some inputs, in particular labour effort, cannot be perfectly contracted over. The importance of this assumption stems from the outcome orientation of a PSM. In Rose-Ackerman’s (1996) terminology, agents motivated by PSM are termed pure altruists. This is different from an action-oriented motivation such as a ‘warm-glow’, under which agents experience an increase in utility just by performing certain actions. With PSM, the nature of the actions performed is irrelevant, all that matters is the effect of the actions. This can give rise to a free-rider problem, as recognized by Gassler (1998):

“I am better off if the poor are better off; I am better off still if you are the one who makes them so.” Gassler (1998, p. 3).

PSM cannot be used to weaken employees’ participation constraints. If an employer attempted to lower employee wages by offering less than the going market wage, no one would apply since they would prefer to see the service provided by someone else. Here, instead, PSM is seen to alter an employee’s incentive constraint. If it can be tapped, a public service motivation leads to the provision of non-contractible effort more cheaply, so that the incentive compatible wage falls. Hence, the non-contractibility of labour effort is critical; with full contractibility, the equivalence between participation and incentive conditions renders considerations based on PSM irrelevant.

The paper relates to the debate on public vs. private provision and establishes relatively precise conditions under which PSM may be a factor favouring government provision of social services. A considerable part of the public administration literature on PSM implicitly assumes that its mere existence implies not only a role for government provision, but also that public sector reform based on implementing management and incentive practices from business can diminish
employee effort based on PSM; see, for example, Gregory (1999). Though in
broad agreement with those arguments, this paper qualifies more precisely when
this can and cannot occur.

The results also provide an explanation for low-powered incentives in bureau-
cracy, as also explored in Banerjee (1997). It will be seen that, formally,
privatization and output-related contracting with employees are not that different.
The paper thus shows that the pursuit of market-oriented reforms, involving higher
powered contracting in the public sector, can actually raise costs. Hence, it may be
optimal not to provide such contracts even though they are feasible. This coincides
with examples drawn from the public administration literature where such
contracting initiatives, in the presence of PSM, were argued to have been
counter-productive. These examples are discussed in Section 6.

The paper proceeds as follows. The next section sets up the information, timing
and primitives of the basic model. Section 3 analyzes service provision incentives
in the private firm and derives the optimal contract for the planner to offer. Section
4 analyzes the optimal set-up of a bureaucracy. Section 5 compares the two
different providers and establishes the conditions under which the bureaucracy
dominates. Section 6 discusses the model’s relationship to case studies and the
literature on public administration and provides a brief conclusion.

2. The model

There exists a benevolent social planner wishing to obtain an exogenously
determined amount of a social service, denoted $g^*$. The level of service provided
is assumed to be fully contractible; the amount of service can be perfectly and
costlessly measured and observed. This assumption distinguishes the model from
Hart et al. (1997) and the great majority of explanations for the existence of
non-profit organizations, as in Weisbrod (1988), Rose-Ackerman (1996) and
Glaeser and Shleifer (1998). The model contrasts the cost of service provision
from two distinct types of institution; a private firm and a ‘bureaucracy’
$. These
will be generically referred to from now on as the ‘provider’. The two providers
are distinct in an important respect. The owner of the private firm is a residual
claimant, in the traditional Alchian and Demsetz (1972) sense; any profits or
losses incurred belong to this owner. In contrast, in a ‘bureaucracy’, no one has

3 This corresponds precisely with the conceptual distinction used in the literature on privatization of
government. This ‘public administration revolution’ emphasized a distinction between the purchaser
and provider or otherwise called the policy and executive functions of government. The changes aim to
preserve the purchaser and policy arms in the public sphere, while examining consequences of
privatizing the provider and executive arms.

4 In practice, contracts on service provision with the private sector are remarkably similar to the very
simple characterization used here; see, for example, Felty and Jones (1998) for a description of actual
contracts between the government purchaser and a managed care provider for health service in the US.
residual claimancy. The formal distinction between the two providers is spelled out further on.

2.1. Primitives, information and timing

All agents are identical in preferences whether working for either type of provider. Furthermore, the production possibilities available in both providers are identical.

The analysis is partial equilibrium so that all prices, other than contracted price for service provision, are taken as exogenous, as well as the level of service provision demanded by the planner, and the alternative wage of workers. With the addition of some complexity, all of these elements could be endogenized but this would add nothing to the basic analysis.

Time proceeds discretely and all agents live forever with discount rate $\beta < 1$ per period. There are no production or consumption interrelations across periods. Thus, all production requires inputs contributed in the current period and any output produced depreciates completely at the end of the period.

Individuals, indexed by $i$, all have the following quasi-linear within-period utility function:

$$u_i = y_i - v(e_i) + \gamma(g) \quad (1)$$

where $e_i = 0$ or 1 denotes $i$’s labour effort, $g$ is the level of service provided per period and $y_i$ is individual $i$’s income (consumption), with $v' > 0$ and $\gamma' \geq 0$; time subscripts are suppressed. The function exhibits the usual feature of separability between labour and consumption, but is non-standard in allowing for the level of $g$ to benefit agents directly. The direct entry of $g$ into utility functions captures individuals’ public service motivation (PSM). As referred to in Section 1, there is considerable behavioural support for PSM. Though some in the public administration literature have contended that PSM is directly linked with working for the government (or a non-profit firm), others have argued that the social motivation need not depend intrinsically on the nature of the provider (Rainey, 1982; Brewer, 1998). In any case, since the model here aims to discover conditions under which PSM cannot be used to motivate workers in a private firm, but can motivate government employees, its existence will be posited equally in employees of either organization. If it is the case that the very ‘publicness’ of the institution itself directly creates PSM, so that preferences change with employer, then this is a direct reason to favour government provision. However, this assertion seems hard to justify, and we consider it no further in the model here.\(^5\)

There exist two types of inputs in production of the service: workers’ labour, the

---

\(^5\)In the final section, a psychological explanation (where preferences change with employer) is contrasted with the more economic one posited in this paper. The advantage of an economic explanation is that it yields testable implications.
aggregate amount of which is denoted \( N = \sum \epsilon_i \), and a vector \((k_1, \ldots, k_K)\) of non-labour inputs, the level of each being described by the \( K\)-dimensional vector denoted \( \vec{k} \). Production of the service, \( g \), is given by the following increasing and concave production function:

\[
g = g(N, \vec{k})
\]

with \( g_1 \geq 0 \) and \( g_i \geq 0 \), \( g_{ii} \leq 0 \), for \( i \in K \).

All workers are able to work elsewhere at an exogenous alternative wage of \( w_a \), with cost of effort \( e_i = 1 \). The utility obtained in alternative work is assumed to be equivalent to the utility from unemployment, thus the pairs \((y = w_a, e_i = 1)\) and \((y = 0, e_i = 0)\) are equal.

As already discussed, with complete contracting over labour effort, participation becomes the worker’s only decision, and a free-rider problem ensues where no type of organization can obtain effort motivated by PSM. Therefore, some contracting incompleteness on the input side is essential here. Incompleteness is introduced by assuming that labour effort is not immediately observable, but can be inferred accurately by the employer after one period, whereas explicit contracts over labour effort cannot be written.\(^6\)

The provider elicits employee effort by an efficiency wage. The provider pays a wage, denoted \( w \), before work effort is observed. If exerting the correct effort, the employee is rehired in the next period once effort is inferred. If shirking, the lack of effort is determined in the next period, the employee is fired, and thereafter loses the wage premium.

The timing of events within a period is as follows:

(i) The planner decides on \( g^* \).

- If contracting with a private provider, the planner specifies a menu of payments it is willing to pay as a function of service provided, denoted \( \{p(g)\} \). As a special case, this includes forcing contracts: \( p(g) > 0 \) if and only if \( g \equiv g^* \).\(^7\)
- If setting up a bureaucracy, the planner hires a ‘bureaucrat,’ and provides the bureaucrat with a budget to produce the service.

(ii) Providers’ decision.

- In the case of a contract, the private provider chooses to accept or reject the contract and, if accepting, in turn, chooses input levels and arranges to hire the workers desired. Workers are paid upfront and other inputs are paid on a fee per service basis.

---

\(^6\)This type of incomplete contract, more formally examined in Macleod and Malcomson (1989, 1998), is commonly used. The relationship between this structure and the free-rider problem is discussed further on in Section 4.

\(^7\)Instead of one contractor I could have equivalently assumed that the contract is awarded via a bidding process between contractors, with fully equivalent results.
In the case of a bureaucracy, the bureaucrat chooses whether to accept or reject employment and the operating budget. If she accepts, she hires government employees, who are paid upfront while other inputs are paid on a fee per service basis.

(iii) Shirking decision.

Workers, in either the private firm or government department, decide whether to work or shirk at their wage.

(iv) Adjustment decision.

The provider, either residual claimant or bureaucrat, estimates the output that will be produced given the input choices in the previous stage. At some cost, the details of which are discussed below, the final level of service can be altered by adjusting the levels of other inputs at this stage.

(v) The service is produced and the period ends.

At the start of the next period, shirkers are revealed to the provider and the above sequence repeats itself.

2.2. Adjustment costs

After the worker’s decision in stage (iii), it is possible for either the owner of the private firm (the residual claimant) or, in the case of government provision, the bureaucrat, to adjust other inputs in stage (iv), and thus affect the level of output produced. Such adjustments may be costly, not only in terms of payments to the extra factors, but also directly to the residual claimant/bureaucrat. Two types of adjustments are of interest: adjustment of labour inputs by hiring additional labour and adjustment of non-labour inputs.

In addition to the extra payments required to hire further inputs, such adjustments will typically involve effort costs to the residual claimant/bureaucrat. In the case of hiring new labour, these are costs associated with finding and instructing appropriate replacements. In adjusting other inputs, these are things such as recontacting suppliers, re-estimating and adjusting production mixes, and so forth. These costs are measured in the direct utility metric by the term $f$, which enters negatively into the utility function.

It is easy to imagine a wide variety of values for $f$ in reality. Some production technologies may simply not allow for adjustment once the worker’s effort has been contributed. In these cases, the value of $f$ will be infinitely high. At the other extreme, it could be that, were a worker to shirk, he or she could be easily replaced by either another worker, with training already in hand, or by another readily available input. In this case, the value of $f$ would be low. Since the parameter $f$ will vary greatly with the application, we take an agnostic approach to it, and present all results conditional upon this variable.
3. The private firm

The private firm is owned by a single residual claimant, who is responsible for all of the firm’s costs and owns all output produced. This residual claimant has identical preferences to any other individual in the model as described in Eq. (1); in particular, he is also potentially motivated by PSM. The residual claimant is not liquidity constrained, i.e. we can think of him being able to finance any non-negative expected value investments. The planner wants the residual claimant, as provider, to produce $g^*$ under the contract $\{p(g)\}$. Assuming the planner can induce $g^*$, which is demonstrated below, the residual claimant solves:

$$\min_{N,k} w_p N + \tilde{r}k, \quad \text{s.t.} \quad g(N,k) \geq g^*$$

where $w_p$ is the incentive compatible wage for the private firm’s workers. Under the maintained assumptions, the solution is well defined, with cost function denoted $c(g^*,w_p,r)$, and the optimal values of choice variables denoted by $N^{p*}$ and $k^{p*}$.

Consider the type of contract offered to the private provider by the planner, $\{p(g)\}$. When designing the contract, the planner is fully aware of the effect it has on the incentives of the residual claimant and his workers, and designs the contract so as to allow the residual claimant to obtain effort as cheaply as possible. Since the alternative to accepting the contract is either unemployment or working at $w_\alpha$, the residual claimant will be induced to participate in the contract provided the expected value of the contract is non-negative.

Though available, a forcing contract will generally not be the lowest cost contract. To see this, consider the two possibly conflicting goals which the planner must balance. The first is to induce the residual claimant to produce the correct level of service, $g^*$. Since output of the service is readily contractible, this could be easily achieved with a forcing contract such as:

$$\{p(g)\} = \begin{cases} Nw_p + \tilde{r}k, & \text{if } g \geq g^* \\ 0, & \text{otherwise} \end{cases}$$  \hspace{1cm} (2)

where $w_p$ is the wage-inducing effort from the residual claimant’s workers.

However, the planner also anticipates the effect of the residual claimant’s
contract on $w_p$. In particular, under the contract in Eq. (2), the residual claimant has a strong incentive to replace inadequate work effort contributed in stage (iii). He is thus unable to credibly commit to leaving inputs unaltered in stage (iv) if one of his workers shirks in stage (iii). If workers are to be motivated by PSM, it is essential that they believe their effort has an impact on the level of service provision. Under (2) they will believe it does not, so a forcing contract will not, generally, be chosen. The exact contract chosen will depend on the residual claimant’s costs of adjusting and the nature of the production function, which we now consider.

3.1. Residual claimant’s incentives to adjust in stage (iv)

Suppose that one worker in stage (iii) has shirked. Without adjustment, output then equals $g(Np^* - 1, \tilde{k}^p)$, and since, in an efficiency wage framework, payments to labour occur before production, the worker has already been paid. Therefore, production costs will still be $Np^*w_p + \tilde{k}^p$. The additional cost of raising output to the level $g^*$ will depend on both the type of adjustments available to the provider and the form of the production function. If it is possible to rehire identical labour, then this will be the cheapest cost method of raising production to $g^*$. In that event, by hiring a replacement worker at a resource cost of $w_p$, the provider still produces $g^* = g(Np^*, \tilde{k}^p)$, so that the additional cost of obtaining $g^*$, given a worker has shirked, is $w_p$. After the period of shirking ends, and the shirker is dismissed, the firm maintains the replacement worker from then on. Alternatively, it may not be possible to adjust labour supply, so that only other inputs can be altered. An earlier version of the paper considered this case which alters the conditions slightly but leaves the qualitative nature of results unaffected.

The reward to adjusting output in stage (iv) depends on $p(g)$. In the case of one worker shirking, the monetary benefit to adjusting is:

$$p(g^*) - p(g(Np^* - 1, \tilde{k}^p))$$

By directly comparing costs and rewards, we obtain a condition under which the residual claimant can credibly commit not to adjust in stage (iv).

**Lemma 1.** The residual claimant will not adjust output when a worker shirks provided:

---

10Under the forcing contract, inadequate production yields zero payment.
11Here we consider the incentives for adjustment when only one worker shirks. This is because, in equilibrium, no one shirks and thus, when considering a worker’s incentive to shirk, the worker uses equilibrium conjectures.
12Alternatively, it may not be possible to adjust labour supply, so that only other inputs can be altered. An earlier version of the paper considered this case which alters the conditions slightly but leaves the qualitative nature of results unaffected.
Proof. If, after one worker shirks, output is adjusted in stage (iv), the residual claimant incurs a cost of $w_p$, as payment to a new worker, and suffers direct adjustment cost of $\phi$. The residual claimant obtains the utility benefit $\gamma(g^*)$ and corresponding payment under the planner’s contract of $p(g^*)$. Alternatively, if not adjusting, no adjustment or extra labour costs are borne. But, since the level of service provided falls, direct utility is $\gamma(g(N_p^*-1,k^p))$, and under the planner’s contract, payments are $p(g(N_p^*-1,k^p))$. These terms give rise to (3) directly. \(\square\)

Note that the residual claimant is also motivated by PSM, so that his or her utility benefit, through $\gamma$, is also considered.

The planner, in designing the contract, wishes to satisfy this condition in order to lower the residual claimant’s wage bill. By inspection of (3) it can be seen that it will be more able to do so the ‘flatter’ it makes the provider’s performance incentives; that is, the smaller the differential $p(g^*) - p(g(N_p^*-1,k^p))$.

### 3.2. Residual claimant’s incentives to produce $g^*$

Flat incentives at the margin are, however, generally incompatible with the residual claimant undertaking to produce $g^*$. In order to ensure $g^*$, there is a minimal level of ‘steepness’ required of the performance contract:

**Lemma 2.** Under the contract $\{p(g)\}$, the residual claimant has incentive to produce output $g^*$ only if:

\[
\gamma(g^*) + p(g^*) - c(g^*,w_p,\tilde{r}) \geq \gamma(g(N_p^*-1,k^p)) + p(g(N_p^*-1,k^p)) - c(g(N_p^*-1,k^p),w_p,\tilde{r})
\]

where $c(g(N_p^*-1,k^p),w_p,\tilde{r})$ denotes the cost function associated with output level $g(N_p^*-1,k^p)$.

**Proof.** Follows directly from substitution into the residual claimant’s utility function. \(\square\)

With insufficient penalty to produce less than $g^*$, the residual claimant optimally chooses to provide a lower level of service, and pockets the cost savings. That is, since $c(g(N_p^*-1,k^p)) < c(g^*)$ this condition requires a sufficiently large differential $p(g^*) - p(g(N_p^*-1,k^p))$ to hold.
3.3. Production vs. non-adjustment incentives

Credibility in the commitment problem in stage (iv) is not always compatible with the incentive condition above. Under some conditions, it is not possible for a private owner to credibly commit to leaving inputs unadjusted when workers shirk, while simultaneously having sufficient incentive to produce \( g^* \). In particular:

**Proposition 1.** If

\[
\phi(g(N^{P*} - 1, k^{P*})) + w + \phi < c(g^*)
\]

then Eq. (3) is incompatible with condition (4).

**Proof.** Immediate by comparing (4) and Eq. (3).\(^{13}\)

Incompatibility between these conditions is more likely: (1) the lower are \( \phi \) and \( w \), the costs of adjustment, and (2) the greater the differential \( c(g^*) - c(g(N^{P*} - 1, k^{P*})) \). This differential represents the cost difference between producing \( g^* \) and the minimal cost method of producing \( g(N^{P*} - 1, k^{P*}) \), and is larger the more complementary are labour and other inputs in the neighbourhood of \( g^* \). When a worker’s effort is highly complementary with other inputs, shirking lowers output by a relatively large amount. Optimally allocating inputs to produce this lower level of service will then cost much less than producing \( g^* \).

Under Eq. (5), the incentives generated by residual claimancy are incompatible with the incentive effect of PSM. Note that this is true for any form of payment contract between the residual claimant and planner.\(^{14}\) The effect on employee wages is examined in the next subsection.

Note also that condition (5) does not depend on the residual claimant’s PSM, since the \( \gamma(\cdot) \) terms cancel out when combining Eqs. (3) and (4). Intuitively, a higher value of \( \gamma \) implies that the planner can design a flatter contract and still motivate the residual claimant to choose the amount \( g^* \). However, this also implies that the residual claimant has greater incentive to adjust output in stage (iv). The effect of reduced steepness is exactly offset by the increase in residual claimant’s incentives to adjust. Therefore, the range of incompatible incentives, given in Eq. (3), does not change. An implication of this is that neither the number of owners, nor their level of concern for the good produced affects the results here.

---

\(^{13}\)Note that \( w \) is treated as exogenous here and determined in the next section. Since the wage determined there is calculated assuming condition (5) holds, it is consistent with its treatment here.

\(^{14}\)When service providers are small, a market price is simply a special case of the contract \( \{p(g)\} \) where \( p(g) - p(g - 1) \) is constant for all \( g \). If service providers are large enough to affect market price, then this increment will decline with their choice of \( g \). Since both of these possibilities are already allowed in the planner setting of a menu \( \{p(g)\} \), I ignore them without loss of generality.
3.4. Wages in the private firm

The lack of contractibility over labour inputs is solved by an efficiency wage (see Shapiro and Stiglitz, 1984). In such a framework, with infinitely lived workers and sufficiently high valuation of the future, it is well known that a finite valued efficiency wage outcome always corresponds to at least one Nash equilibrium of the repeated game between workers and the firm, under appropriate subgame perfect strategies. I proceed here by simply determining the efficiency wage directly from workers’ incentive compatibility conditions, treating implicitly the precise strategies played by the worker and firm in their repeated interaction.\cite{footnote15}

**Lemma 3.** Under condition (5) the private firm’s incentive compatible wage, $w_p$, is:

$$w_p = w_a + \frac{1 - \beta}{\beta} [v(1) - v(0)]$$  \hfill (6)

**Proof.** The proof is in Appendix A.

This wage reflects a standard efficiency wage structure. Due to the possibility of shirking, a premium, reflecting the effort cost, $v(1) - v(0)$, weighted by the discount factor, must be paid over the alternative wage to induce effort.

4. Government production

We model government in a stylized manner. At the heart of the definition of government bureaucracy, and crucial for this analysis, is the non-existence of an individual or group with a residual claim on profits. Thus, a ‘bureaucrat’ is hired by the planner to provide $g^*$, and allocated a budget with which to do so. The bureaucrat, however, has no claim on the budget for personal use and, although motivated by concern for the level of service like all agents, has no concern for the level of operating profit or loss within her department.\cite{footnote16}

The collection of contracts, implicit and explicit, defines the government bureaucracy here. In her employment contract with the planner, the bureaucrat is given a budget, $b$. The bureaucrat fills an analogous role to that played by the residual claimant in the private firm. The bureaucrat writes implicit contracts with individuals to provide effort; these individuals are government workers. Since she

\footnotetext[15]{It should be noted that this is not the only equilibrium possible; the reader should see Macleod and Malcolmson (1989) for a more detailed analysis. In particular, there may exist degenerate equilibria, depending on the specification of beliefs, which allow for no possibility of production.}

\footnotetext[16]{Of course, the planner could make such concerns part of her contract but will choose not to do so, under the conditions we establish.}
faces the same problem in contracting over effort as did the residual claimant, these contracts will also be of an efficiency wage type. They specify a wage and rehiring in subsequent periods as a function of worker effort, as in the private firm. Were the bureaucrat truly unconstrained in her ability to go over budget, she would simply choose to provide an infinite amount of service, or at least an amount in excess of $g^*$, in accordance with her utility maximization. However, the bureaucrat must not plan to spend more than the budget allocated by the planner; that is, for any chosen $g$, $c(g) \leq b$. The bureaucrat’s problem is thus to maximize her utility subject to the budget allocated by the planner.

It is then easy for the planner to induce $g^*$. Denoting the incentive compatible wage for government workers by $w^*_b$, this is done by the planner allocating the bureaucrat a budget of $b^*$ such that the solution to:

$$
\max_{N,k} g(N,k), \quad \text{s.t. } w_N N + \bar{r}k \leq b^*
$$

equals $g^*$. Note that, since there is no effort cost to the bureaucrat, her pay can be set to zero, which is identical to the implicit wage that a residual claimant receives when accepting a contract from the planner in Section 3.\(^{17}\)

Under such a budget, the solution values are well defined and are denoted $N^{g^*}$ and $k^{g^*}$, with $g(N^{g^*},k^{g^*}) = g^*$. The budget is set so that it is just sufficient to finance expenditure on inputs required to produce $g^*$. Lower levels of service are feasible but would lower her utility, hence in a bureaucracy there is no need for strong incentives.

Note that the distinctive feature of government production (no residual claimant) could be altered by an appropriate contract between the planner and the bureaucrat. It would be possible to mimic some elements of residual claimancy by tying some part of the bureaucrat’s payments to service levels. In the extreme, this is nothing other than ‘privatizing’ the bureaucrat’s department, and replacing the bureaucrat with a residual claimant. However, under the conditions that will be established, costs are strictly lower under government provision, so that the bureaucratic structure outlined above dominates more high-powered incentives. We return to this point in Section 6.

### 4.1. Stage (iv) adjustment in the bureaucracy

Since there is no residual claimant, the question arises as to who would undertake adjustment in stage (iv) were some workers to shirk in stage (iii). If the

\(^{17}\)In steady state, since there is no turnover and thus no possibility of obtaining efficiency wage employment, this is the correct specification of alternative utility to a bureaucrat or a residual claimant. In a more realistic model, with turnover, things are more complex but qualitatively the same. There the value of unemployment would rise to reflect the probability of transitioning into an efficiency wage job, but results would be unaffected.
budget constraint facing the bureaucrat binds strictly, then bureaucratic commitment to non-adjustment in stage (iv) is immediate. However, this seems an underhanded comparison with the residual claimant, since it effectively assumes away the issue of bureaucratic incentives to adjust. For a fair comparison between bureaucracy and the private firm, capacity to adjust inputs in stage (iv) should be identical. Thus, it is assumed that the bureaucrat does have authority to use state funds to adjust output in stage (iv), were a government worker to have shirked in stage (iii). This means that the only difference between the bureaucrat and the residual claimant is that implied by the definition of residual claimancy; both face the same production constraints, both have the capacity to adjust production in stage (iv), and both make decisions based on their own utility calculations, but only the residual claimant appropriates any monetary gain or loss.

Consider now the bureaucrat’s incentive to make adjustments when one worker shirks in stage (iii). Though the bureaucrat is not directly rewarded for such adjustment, she may still do so if her private benefits due to PSM are great enough. However, with \( \phi \) sufficiently large, no adjustment will be made:

\[
\gamma(g^*) - \phi < \gamma(g(N^{g*} - 1, k^{g*}))
\]

Under condition (8), government workers in a bureaucracy know that, were they to shirk in stage (iii), the level of service provided would fall. Thus, their effort in production matters, and their PSM affects their incentive compatible wages.

4.2. Wages paid to government workers

Wages paid to government workers are determined analogously to those in the private firm, except that, when condition (8) holds, workers know that their shirking will affect the level of service. This implies that a worker’s PSM enters into their incentive compatible wage:

**Lemma 4.** Under condition (8) the incentive compatible wage for government workers, \( w_g \), is:

\[
w_g = w_p + \frac{1 - \beta}{\beta} \left[ \gamma(g(N^{g*} - 1, k^{g*})) - \gamma(g^*) + \nu(1) - \nu(0) \right]
\]

**Proof.** The proof is in Appendix A.

The term \( \gamma(g(N^{g*} - 1, k^{g*})) - \gamma(g^*) \), in contrast with Eq. (6), reflects the worker’s impact on the service level. Direct comparison of this wage with \( w_p \) from (6), immediately reveals that \( w_g < w_p \). Note that the free-riding problem, referred

---

\[^{18}\text{Since } g \text{ is verifiable, this could always be formally written into the bureaucrat’s operating guidelines to ensure that the bureaucrat did not produce more than } g^*.\]
to in the introduction, need not arise here. A worker’s efficiency wage is lowered by PSM below \( w_p \), but provided it exceeds \( w_p \) (the wage in alternative employment), all workers will prefer to obtain employment in a government job. Thus, provided that:

\[
v(1) - v(0) > \gamma(g^*) - \gamma(g(N^* - 1\bar{k}^*))
\]

participation constraints are satisfied and the free-rider problem does not ensue. Intuitively, provided a worker’s PSM is less than their personal gain from shirking, their incentive compatible wage will strictly exceed the wage satisfying participation. All workers would rather work for a provider than work in an alternative occupation. I assume that (10) holds in what follows.

To summarize, under condition (8), the bureaucrat acting under the budget constraint in (7), both chooses \( g^* \) and credibly commits not to adjust when a worker shirks. The difference between the bureaucrat and residual claimant arises because operating profit plays no part in the bureaucrat’s decision. Thus, in providing incentives to produce \( g^* \), the planner is able to utilize the bureaucrat’s own concern for output when setting the contract and thus, without extra production incentives, does not reduce the bureaucrat’s credibility not to adjust in stage (iv). In the private firm, in contrast, even though the residual claimant cares as much for output as does the bureaucrat, his concern for his own income also enters. Under condition (5), it is this concern which makes production incentives incompatible with incentives not to adjust in stage (iv). Since \( w_p \) is calculated under (5), and \( w_b \) is calculated under (8), it only remains to compare these conditions to obtain the situations under which \( w_b < w_p \).

5. A comparison of both providers

Since \( g^* > g(N^* - 1\bar{k}^*) \), we can immediately derive conditions under which government provision has a cost advantage over a private firm:

**Proposition 2.** If

\[
c(g^*) - c(g(N^* - 1\bar{k}^*)) - w_p > \phi > \gamma(g^*) - \gamma(g(N^* - 1\bar{k}^*))
\]

then only a bureaucracy can pay workers a lower incentive compatible wage due to their PSM and \( w_b < w_p \).

**Proof.** The second inequality implies that (8) holds, so that the bureaucrat’s commitment in stage (iv) is credible. Thus, the costs of bureaucratic provision are \( c(g^*, w_b, \bar{r}) \), where \( w_b \) solves (9). The first inequality implies that condition (5) holds, so that there does not exist a payment contract which allows for a credible commitment in stage (iv) in the private firm. The private contractor must therefore
pay \( w_p \) so that costs of provision are \( c(g^*, w_p, r) \). A direct comparison of (6) and (9) makes clear that \( w_p > w_s \). □

The two different parts of expression (11) depict the principal trade-offs involved. The latter inequality represents the difference in personal valuation of the service provided when adjustment is made. This depends on the production technology, but also on the intensity of a worker’s public service motivation. In contrast, the left-hand side represents the difference in minimal costs of producing \( g(N^{p*} - 1, \tilde{k}^{p*}) \) instead of \( g^* \). This is independent of PSM and derives directly from the production technology.\(^{19}\)

The difference in factors driving the two inequalities can be seen most clearly by considering an extreme case, such as highly interdependent Leontief production. In a Leontief production function, where every worker’s input is critical, \( g(N^{p*} - 1, \tilde{k}^{p*}) = 0 \). In this interpretation, the left-hand side of (11) represents the difference in costs between producing the total level of service and none at all, minus the wage. In dollar terms, this could be very large. The far right-hand term represents the dollar equivalent utility fall suffered by one worker when the service is not provided. This could also be large but, of course, does not bear any necessary relationship with the cost of the service. In general, however, it would be surprising to find this larger than the cost of service provision, especially when considering a service that is a public good. Thus, examples of services that fit these parameters are those usually emphasized in discussions of PSM. These are typically ‘care’ services where providers have a strong commitment to the welfare of recipients such as, health care, housing, social work, some forms of education, social protection agencies, etc. The critical constraint implied by (11) is that the cost of social service provision exceeds the utility benefit accruing to the provider (be it residual claimant or planner) from providing it.

The existence of other inputs in the production of the service, and their complementarity with effort, is also critical. Without the non-labour terms in production, or if production were separable in labour and non-labour components, optimally applying labour up to the equality between marginal benefits and costs will imply that the term on the left-hand side of (11) generally equals zero. With complementarity in production at the margin, however, the possibility arises for a relatively large difference between the cheapest cost method of producing \( g(N^{p*} - 1, \tilde{k}^{p*}) \) and the cost of \( g^* \), which, as shown in Section 3, is necessary for a private firm to be unable to obtain effort motivated by PSM.

The model yields predictions as to when one would expect only a bureaucracy to obtain labour effort that is motivated by PSM. If PSM is too great, condition (11) will fail since government workers will understand that, even if they do not

\(^{19}\)Note that these conditions do not conflict with the earlier necessary condition for avoiding the free-rider problem, condition (10), since that also implies a lower bound on the far right-hand side term of (11).
work, the bureaucrat is motivated enough to cover for their reduced effort by her own PSM. An example where there is likely to be PSM but where (11) is likely to fail, is the case of dike provision analyzed by Cornes (1993). The strength of a dike can be well characterized by a Leontief technology so that the complementarity requirement is met, and aggregate costs are high. However, personal costs to a dike failing are likely to be higher still so that the right-hand inequality will generally fail to hold. In the case of dikes then, the analysis here suggests no reason to favour public over private providers.\textsuperscript{20} The value of $\phi$, on the other hand, can be neither too large, in which case the private firm can credibly commit to not adjusting, nor too small, in which case, even a bureaucratic manager may have incentives to adjust.

Another interpretation of these results pertains to the effects of increased performance-based contracting within an organization.\textsuperscript{21} As argued in Section 4, the planner could write a higher powered, i.e. output-based contract, with the bureaucrat without having to go so far as privatizing the department. In that case, under condition (11), the model predicts that the introduction of such high-powered incentives for the bureaucrat crowds out the PSM motivated contributions of government workers and raises costs, providing a partial explanation for the observed lack of such contracts in government relative to private firms.\textsuperscript{22} Of course, the well-known reasons for writing such performance-based contracts in the first place (for example, bureaucratic moral hazard) are given no role in this model, so the model is not a normative argument for either government provision, or low-powered incentives. Instead, it suggests considerations that may be important in the privatization debate, and which have received little scrutiny from economists thus far.

6. Discussion and examples

The clearest examples of the considerations here arise when the PSM of individuals is not confounded by other incentives. The analysis of volunteers is thus instructive since financial motivation is less likely to be present. Indeed the

\textsuperscript{20}A further consideration in this case is that condition (10) is likely to fail, so the free-rider problem may ensue. A similar problem has been analyzed by Engers and Gans (1997) in the case of low-powered incentives for referees. There, in contrast, the free-rider problem is solved by there being a cost to delay after refusing participation.

\textsuperscript{21}See Banerjee (1997) for an explanation based on Holmstrom and Milgrom (1991)-type agency considerations.

\textsuperscript{22}See Scott and Falcone (1998) for a survey of the comparative literature.
case studies cited below suggest the prevalence of PSM in volunteers. In estimating the relevance of the concerns in this paper, the empirical question, then, is does the implementation of higher powered incentives at the top of an organization crowd out voluntary contributions at lower levels? There is some evidence of precisely this occurring.

Simpson (1996) describes a situation where the contribution of voluntary labour to community-based fire-fighting organizations was compromised by cost-reducing considerations implemented at the locality level. In a desire to increase outside investment within these localities, the local governments of Clinton county in the north-east of New York state, attempted to reduce local expenditures on fire services by encouraging the expansion of voluntary services. The reasoning was that the consequent lowering of taxes and provision of inexpensive emergency services would encourage outside investors to locate in the region. However, as Simpson notes, when voluntary fire department service became perceived as a tax on voluntary members which served to reduce payments necessary for career fire-fighters, the voluntary process started to lose its legitimacy and risked being deserted by the volunteers. He argues that the trend of a shortfall in volunteers reflects this perception. In terms of the model, the attempt to economize on paid inputs by using volunteers severed the link that volunteers viewed between their effort and the level of service. In that case, though volunteers presumably still felt a PSM, they no longer felt a desire to contribute effort on its behalf.23

A similar unintended effect of contracting on volunteers was reported when non-profit community service providers in New Zealand experienced an increase in the formalization of their contracts with service purchasers.24 The impetus for such changes came from public service initiatives, in particular a 1991 public finance act aimed at increasing efficiency and mimicking effective business practices, such as output-based contracting. The author describes the effects of these initiatives on volunteers:

23 Although this is not an example in government it is useful because the lack of other forms of payment allowed the reduction in effort motivated by PSM to be clearly observed through the fall in volunteers. In public administration, the effect is likely to be much more subtle because employees will typically still contribute work. The literature on performance pay in the public service has documented considerable difficulties in its operation; see, in particular, Kellough and Lu (1993) and Ingraham (1993). However, studies of the effects of performance pay are not that helpful in assessing the implications of this paper since performance pay or increased contracting is rarely implemented in isolation. Instead, such changes usually coincide with reduced funding for social services and large-scale organizational change. Thus, case studies documenting lower morale in service providers, for example Motenko et al. (1995), or an erosion of standards such as Gregory (1999), do not disentangle the effects of lower funding, organizational change, and difficulties in contracting from the crowding out of PSM suggested here.

24 This account is taken from Nowland-Foreman (1998).
“a pure competitive-tendering model risked the volunteer contribution on which services rely . . . A legalistic contract approach can be useful for clarifying expectations but risks reducing the ‘gift relationship’ of the volunteer to a transaction and consequently undermining the intrinsic motivation of volunteers members and other supporters” . . . (p. 117).

He goes on to argue that though there were short-term gains in increased utilization of volunteers to meet contracted objectives, this risked a longer term disenchantment and erosion of volunteer contributions when they came to perceive that instead of adding value, they were substituting for government actions.

Though not examples of privatization, the very motivation for these changes has consciously been the mimicking of the incentive structures in the private sector and these concerns about ‘crowding out’ of civil-minded activities are readily understandable in the context of the model used here.

This precise theme of contracting serving to undermine voluntary contributions is echoed time and again in response to increased contracting initiatives, both in the implementation of higher powered contracts within government bureaucracies, and in the extension of contracting with not-for-profit service providing agencies. The following quote from LeGrand (1998) summarizes these arguments:

“In particular, a publicly provided service will not be one where everyone is out to feather their own nests, but one where individuals work to meet the interests of their clients. Indeed, the introduction of private provision and market mechanisms may worsen the situation. For this would increase the rewards to self-interested behaviour, thus encouraging it to develop at the expense of more altruistic concerns . . .” (p. 418).

In the analyses of contracting within government, the risks are often discussed in terms of impacting upon standards of ethical probity, i.e. undermining the virtues of public merit and civil service. In non-profit organizations, the argument is made that enhanced contracting compromises the service ethic on which non-profits are founded. In either case, it is tempting to interpret these examples as situations where the very nature of actors’ preferences have changed by the implementation of contracts. Indeed, this is often the interpretation favoured by writers in public administration, where it is argued that the very inability of economists to account for such changes in preferences, or to even allow for non-selfish motivation, renders traditional agency-based approaches of little relevance; see, for example, Rainey (1991) (Chapter 1), or Worsham et al.

---

25See, for example, Gregory (1999).
26See, for example, Ryan (1999).
(1997). The model here counters that assertion by presenting an alternative explanation of the same phenomena based on purely economic reasoning, while, importantly, assuming that preferences do not change.

An advantage of the economic approach is that it yields predictions as to when such contracting is likely to have ill effects, from (11). To recap, the incentives created by residual claimancy, or high-powered contracting, are more likely to be incompatible with employee effort motivated by PSM when:

1. Employee effort is complementary with other inputs.
2. Employees’ PSM is not too great.
3. The costs of adjustment, $\phi$, are neither too large nor too small.

All of these implications are potentially testable, and do not coincide with the predictions of other theories explaining low-powered incentives, or government provision.

I know of no evidence relating to the first two implications in the literature. However, if such considerations, as opposed to more psychologically based ones, are important, we may expect to find examples where a residual claimant (or analogously an agent with high-powered incentives) undertakes actions to increase the size of $\phi$. In line with the third implication above, the agent’s aim would be to ensure that the first inequality in (11) fails, so that their commitment not to adjust inputs could be maintained even when they have high-powered incentives. A real-world analogue of this would be for the residual claimant to maintain an arm’s length relationship, perhaps by credibly undertaking not to audit and evaluate employee performance until after services are rendered. If there is no way even to know of their employees’ performance, then this would correspond to an infinitely high value of $\phi$, and commitment would be immediate.28

Though there is no formal empirical documentation of such effects in the management literature, nor in the analysis of public-sector contracting, numerous writers on public administration have emphasized the importance of goal responsibility and autonomy. That is, the capacity of public servants to pursue tasks

---

27Most of the criticism of neo-classical economics in this tradition does, however, display a rather uninformed critique of extremely simplified models. More thoughtful critiques from the public administration perspective can be found in Holtham and Kay (1994) and Gassler (1998). Frey (1997a) formally models these psychologically based perspectives by allowing for preferences to change endogenously.

28An earlier version of the paper looked at the contracting requirements necessary for a residual claimant to create an arm’s length relationship, and thus replicate the PSM of a bureaucracy in a private firm. It was shown there that this could only happen when contracting over input levels, or expenditures was complete. Even reputation-based incomplete contracting over inputs could not lower the private firm’s costs to that of a bureaucracy. This version is available upon request.
from start to finish, and to be relatively free from interference in the achievement of these. This is one of the critical features in effective government organizations emphasized in Denhardt (1993) (in particular, the cases studies and discussion in pp. 133–170). Such emphasis on maximal delegation of responsibility and authority to workers is also found in Wilson (1989), Gold (1982) and Rainey and Steinbauer (1999). This is consistent with the implications of the present model though more careful evaluation of the model’s other predictions awaits further research.

Acknowledgements

I have benefited greatly from the comments of two anonymous referees and Roger Gordon. The work grew out of discussions and correspondence with Joshua Gans to whom I am greatly indebted. I have also benefited from discussions with Siwan Anderson, Robin Boadway, Jim Minifie, Pierre Pestieau and Tanguy van Ypersele and from comments of participants at the Queen’s Public Finance workshop. I am especially grateful to Dan Usher, Ashok Kotwal, Pieter Ruys and Jano Zabojnik for extensive comments. John Alford helped greatly in alerting me to relevant work in the public administration literature and Renu Mandhane provided outstanding research assistance which was supported by SSHRC.

Appendix A

A.1. Derivation of $w_p$:

Define the following value functions: $V^h$ is the expected lifetime value of being honest, $V^s$ is the expected lifetime value from shirking, $V^u$ is the expected lifetime value of being employed in the alternative job until death. For the for-profit firm under condition (5), the level of service produced is unaltered by the worker’s shirking. Thus:

$$V^h = w_p + \gamma(g^*) - v(1) + \beta V^h$$
$$V^s = w_p + \gamma(g^*) - v(0) + \beta V^u$$
$$V^u = w_a + \gamma(g^*) - v(1) + \beta V^a$$

Thus, at the binding no shirking condition, $V^h = V^u$, we have four equations which can be easily solved for all values of $V$ and $w_p$, yielding:

$$w_p = w_a + \frac{1 - \beta}{\beta} [v(1) - v(0)]$$
A.2. Derivation of $w^*_g$:

For the non-profit firm, under condition (8) no adjustment of other inputs will be made in the event of shirking, so that shirking has real effects. Thus, the value functions are:

\[ V^h = w^*_g + \gamma(g^*) - v(1) + \beta V^h \]
\[ V^s = w^*_g + \gamma(N^* - 1, k^*) - v(0) + \beta V^u \]
\[ V^u = w^*_u + \gamma(g^*) - v(1) + \beta V^u \]

By once again imposing the incentive compatibility condition, $V^h = V^s$, we can solve for the four endogenous variables, including the wage, $w^*_u$. This simplifies the wage expression to:

\[ w^*_g = w^*_u + \frac{1 - \beta}{\beta} \left[ \gamma(N^* - 1, k^*) - \gamma(g^*) + v(1) - v(0) \right] \]

References


