The Limits Of Bureaucratic Efficiency

Canice Prendergast*
University of Chicago & NBER

Third Draft: May 9, 2001

Abstract

Bureaucrats typically intermediate between a principal and a consumer, by diagnosing benefits for the consumer. This paper argues that bureaucratic efficiency is limited by the fact that the decisions made by bureaucrats involve rents to consumers. This means that a primary means of oversight, namely, using consumers to complain about incorrect decisions, can become ineffective. This has two implications for a bureaucracy. First, oversight becomes more difficult as customers cannot be relied upon to point out bureaucratic error. Second, it gives bureaucrats an incentive to accede to consumer demands simply to avoid a complaint. I show that when this second effect is important, bureaucracies (efficiently) respond in the following ways: (i) they ignore legitimate consumer complaints, especially those aimed at incompetent bureaucrats, (ii) they monitor more in situations where it is not needed, (iii) they correct fewer errors than in non-bureaucratic situations, (iv) they delay decision-making "too long", and (v) oversight is biased against consumers.

This paper also shows how the need for bureaucrats depends on how their allocations are priced to consumers. The primary implication of this section is that observed bureaucracies are always inefficient: the features that make bureaucrats more efficient also make them unnecessary. To phrase this another way, when bureaucracies work well, consumer choice works even better. By contrast, when bureaucratic choice works inefficiently, consumer choice works even worse. Thus bureaucratic organizations appear to work less well than those where consumers have more choice, yet here this is no fault of the bureaucracy.

*I am grateful to Kent Daniel, Lars Stole, and seminar participants at Harvard University, University of Michigan, Bristol University, and the Royal Economic Society, Durham, for helpful comments. I would also like to thank the NSF and the University of Chicago for generous support. Any errors are my own.

1 Introduction

Bureaucrats pervade economic life. They approve our medical procedures, process our credit card enquiries, decide whether to arrest and incarcerate us, issue our licenses, approve our immigration status, schedule our appointments, and so on. Arguably most economic interactions that we engage in involve not the canonical buyer-seller relationship of economic theory, but are instead affected by some intermediary. The objective of this paper is to better understand agency issues that affect bureaucratic decision-making and to identify the constraints that make efficiency difficult to attain.

It is hard to find much good that is said about bureaucracies, both private and public. The IRS and INS are regularly vilified in the press, and the practices of health insurance companies' bureaucracies and police officers fare little better. This is reflected in the pejorative terms for bureaucrats (beancounters, penpushers, and so on) that pervade such descriptions. The typical perception of a bureaucracy has some of the following features. Standards of consumer service are low. They are largely unresponsive to customer complaints. Their decisions are rarely overturned. They are predisposed to turning down consumer requests. They take forever to come to decisions. Finally, they appear to be governed by rules (perhaps the defining characteristic of a bureaucracy) rather than using their discretion in the appropriate way. This paper offers a model of bureaucracies that yields these outcomes as the optimal resolution of agency problems. Perhaps most importantly, bureaucrats are used only when they exhibit these "inefficiencies": ironically, the factors which lead bureaucrats to be more efficient also render them unnecessary.

Bureaucrats typically mediate between a principal and a consumer, by diagnosing whether a consumer should receive some benefit. In this paper, a bureaucrat is defined as an individual who has control over the allocation of a benefit to another, a customer, where control of the allocation derives from private information that she holds over its optimal use.[1] I argue that agency problems from bureaucracy arise for two reasons. First, the decisions made by bureaucrats *optimally* involve *ex post* rents to consumers. (For example, it should matter to a patient that he is approved for a medical procedure, to an applicant that he be given a green card, or to a suspect that he not be arrested.) Of course, many goods allocated by other mechanisms involve ex post rents for consumers: this is what we call consumer surplus. The second characteristic that leads to bureaucratic problems is that although consumers are interested parties, they cannot be trusted to allocate the benefits. To give a ridiculous (but relevant) example, it is the rare suspect who would arrest himself if given a choice between that and setting himself free. I illustrate that these features

[1] For example, an arrest should be made by a police officer only if there is a suspicion of guilt, something that is only immediately apparent to the officer herself. By contrast, a sales assistant or an auctioneer would not qualify by this definition as the decision to transfer the good depends only on the price paid, which is easily observed.

simultaneously explain both why bureaucrats are used and why bureaucratic agency problems arise. Thus, some bureaucratic practices arise from the nature of the goods that bureaucrats allocate, rather than the inherent inefficiencies of bureaucrats *per se*.

This paper is concerned with inducing efficient performance from bureaucrats. Bureaucrats are rarely offered pay-for-performance based on easily available outcome measures.[2] Nor are they offered rewards based only on the allocations they propose.[3] Instead, the primary way of controlling the behavior of bureaucrats is by costly investigation of the details of cases. These investigations, whether formalized in the various commissions, tribunals and special investigations of the public sector, or through more informal oversight, are the way bureaucrats are typically monitored. But such oversight is not randomly assigned (as in Becker, 1968), nor should it be. Instead, investigations are targeted to cases where a mistake is likely to have been made, and are typically triggered by particular signals of a mistake or malfeasance.[4]

Perhaps the ubiquitous signal that focuses attention on a case is a *consumer complaint*, which is used because consumers have legitimate information on the correct allocation of goods, and raising flags helps superiors to intervene. This is the stuff of bureaucracies, where the best that managers can hope to do is to "step in when complaints are heard or crises erupt" (Wilson, 1989, p.175). Complaints are, of course, also used in most non-bureaucratic organizations. For instance, poor service in a store will often result in a request to "see the manager". I argue here that complaints mechanisms have particular problems in bureaucratic settings. The reason for this is that typically the consumers of bureaucracies have preferences that do not correspond with social welfare, and can obtain benefits from being (inefficiently) allocated rents by the bureaucrat. For example, patients generally wish to be approved for medical procedures even when it is not efficient for the procedure

[2] Firms generally use the following hierarchy when attempting to relate employee rewards to performance. First, they consider basing pay and promotion prospects on easily observed measures of performance. If such easily observed measures of performance correspond closely to the vector of objectives of their employers, this generally becomes the preferred mode of reward. However, writing contracts on such measures is likely to be counterproductive in many occupations as a result of multitasking concerns (Holmstrom and Milgrom, 1992), such as the examples described in Prendergast, 1999, where rewarding individuals on certain measures results in excessive focus on these measures to the detriment of unmeasured factors. Most bureaucratic jobs are in this class. The next possibility typically considered is to monitor processes (or "inputs" in the terminology of agency theory). In some situations, this approach works well, such as where a driver's license can be renewed after a sight test and evidence of a reasonable prior driving record. Yet in other situations, excessive reliance on processes becomes problematic. For instance, deciding on the criteria by which police officers arrest a suspect is likely to be too difficult, given the enormous variety of situations that they face. For occupations where neither easily available performance measures nor process monitoring can guarantee that employees act appropriately, some other means of controlling malfeasance is necessary. These organizations, which Wilson calls "coping organizations", have the most difficult agency problems to resolve, and are the subject of this paper.

[3] For instance, police officers are not rewarded when they arrest someone, nor are benefit officials rewarded whenever they deny benefits to a consumer.

[4] For example, "managers of police patrols, like managers of operators in any coping organization, try to achieve compliance by attending to alarms - periodic signals that something has gone wrong" (Wilson, 1989, p.175).

to be done, and suspects do not want to be arrested.

Problems then arises for two reasons. First, if a consumer is mistakenly given rents, he will not complain. This implies that bureaucratic investigations are less precisely focused because consumers cannot be trusted to reveal that an error has been made. My main interest here is in a second problem with complaints mechanisms for bureaucrats, namely, the harmful incentives that they imply. Bureaucrats are well aware that their performance is under the spotlight when complaints are made against them. Not surprisingly, this means that from the bureaucrat's perspective, "all that matters is that there are not 'too many' complaints" (Wilson, p.175). This implies that she has an incentive to give customers what they want, even when it is not socially efficient, simply to avoid the possibility of a complaint. For example, a police officer could choose not to arrest someone to avoid the possibility of a wrongful arrest complaint or the possibility that she has used excessive force. Similarly, an INS official could allow an unqualified candidate to enter the country rather than avoid the type of case reported in the New York Times, 2000, where the officials were accused of racism. Finally, consider the effect of the recent increases in oversight of the IRS. This has resulted in "a sharp roll-off in tax investigations as auditors, fearing for their bureaucratic lives, proceed timidly..[as]..tax collectors are too worried about their jobs to be aggressive" (Star Tribune, 2000).[5]

Specifically, this problem gives rise to a *truth-telling* condition: what practices must be used to induce the bureaucrat to honestly deny benefits to the consumer? I argue that many practices of bureaucracies exist to overcome the temptation to capitulate to consumers simply to avoid complaints. Much of the paper is to show how judicious use of monitoring propensities, penalties, and timing of decisions can improve the decisions made by the bureaucrat, even when faced by this fear of oversight triggered by complaint.

The central concern of the paper is bureaucratic oversight. The bureaucracy's optimal policies depend on the threat that a complaint imposes on the official when oversight is set optimally to correct bureaucratic error. If the bureaucrat has little fear that she will be found to be wrong, I show that she can be induced to report honestly and exert effort with no distortion in monitoring propensities. This is the situation where the truth-telling constraint does not affect organizational practice. On the other hand, when bureaucrats feel threatened by complaints and investigations, the truth-telling constraint is violated when oversight is set at the efficient level, because the bureaucrat prefers to give in to the consumer and thus reduce the likelihood of investigation. In that case, I show that the following policies will be used. First, bureaucracies become less

[5] See www.trac.syr.edu for details of the recent reduction in audit rates by the IRS.

responsive to complaints, even though complaints reveal that bureaucratic error has occurred. Second, bureaucracies increase monitoring in the absence of a complaint beyond its efficient level. Thus they have more oversight in cases where there is little need for it. This apparently inefficient way of monitoring is used to induce the bureaucrat to deny benefits to the consumer and run the risk of a complaint. This basic insight is analyzed in Section 3. I also show here that when these practices are used, incompetent bureaucrats are monitored less frequently when complaints are made than their more able counterparts, even though they make more mistakes. This is counter to standard economic logic, where the superior oversees more when he thinks mistakes are more likely. The reason for this is that the incompetent are more threatened by an investigation and, thus, must be particularly shielded from complaints if they are to deny benefits to consumers and run the risk of a complaint.[6]

Section 4 considers other implications of these bureaucratic constraints. First, bureaucratic oversight is biased against consumers. This arises because, (i) by ignoring complaints, superiors intervene too little if the consumer is incorrectly denied benefits, and (ii) by over-scrutinizing cases with no complaint, they intervene too much when the consumer is (sometimes incorrectly) given the asset. This results in "too few" denials being overturned and "too many" approvals being overturned. A second implication is that when the cost of monitoring is quadratic, this policy of ignoring some complaints and over-monitoring routine cases results in (i) a higher average probability of monitoring than in non-bureaucratic settings, but (ii) fewer mistakes corrected. Thus, this section offers a theory of bureaucracy which intervenes more frequently, but does so in such a haphazard fashion that it corrects fewer errors.

A third implication concerns the speed at which bureaucratic decisions are made. One way to make bureaucrats less worried about investigations is for them to be more certain before they make a decision. I show that when bureaucrats worry enough about the prospect of a customer complaint, decision-making is delayed more than is technologically efficient in order to make bureaucrats more sure that they are making the correct decision. In this way, they become more likely to reveal their findings truthfully rather than capitulate to the desires of consumers.

The results of the paper are predicated on the assumption that bureaucrats wish to avoid investigation of their actions. Section 5 formalizes this assumption by addressing a career concerns setting where the actions of the bureaucrat become more observable upon an investigation: this assumption yields the ex ante desire to avoid investigations, and the bureaucratic outcomes that arise.

[6]See Freibel and Raith, 2001, for other work on how limiting communication in firms may improve incentives.

Bureaucratic allocation therefore suffers from a series of inefficiencies. But if bureaucrats are so inefficient, why are they used? Section 6 describes how the need for bureaucrats depends on the pricing of benefits to consumers. I make two claims in this section. First, bureaucrats should be used only when their decisions are not priced; when benefits are priced appropriately, consumers should be allowed to choose for themselves. Bureaucracy should then be limited to cases where such underpricing is optimal. I focus on two such cases; those involving insurance and incentives. Second, efficient bureaucracies should never be observed, because the condition that makes bureaucracies efficient also renders them unnecessary! In effect, whenever bureaucracies functions effectively, consumer choice works better.

2 The Basic Model

An allocation A must be made to a consumer, where A can take on a value of 0 or 1. The social surplus from the allocation depends on a parameter α and is given by

$$S(A;\alpha) = \begin{cases} 1 & \text{if } A = \alpha, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus social surplus is positive only if A is properly matched to the underlying environment, α. The true value of α is unknown and can take two values $\alpha = 1$, or $\alpha = 0$. In this section, I assume that each state occurs with equal probability.

Information and Objectives There are three actors in this model, a principal, an agent (or bureaucrat), and a consumer. First, the agent collects information on α; she observes α_a which is correct with probability $q \geq \frac{1}{2}$, where with probability $1-q$, she observes $\alpha_a \neq \alpha$. The precision of the agent's estimate depends on her unobserved effort decision, where she chooses between high effort $e = 1$ or low effort $e = 0$, where high effort has a disutility of d. Let $q(e)$ be the precision, where $q(1) > q(0)$. The agent's objective is to maximize wages minus effort costs. I assume throughout that in the absence of agency issues, it is efficient for the agent to exert effort.

Second, the consumer observes α.[7] Rents earned by the customer play a central role in the ability of bureaucracies to function effectively. Accordingly, let $V(A,\alpha)$ be the utility obtained by the consumer if his type is truly α and the allocation is A.

[7] The case where consumers are imperfectly informed is explored in Prendergast, forthcoming: the qualitative effects remain unchanged.

The principal is uninformed unless he carries out an investigation. To model a role for investigations, I assume that the principal chooses a probability of observing a signal on the true state of nature, α, at some cost. Specifically, the principal chooses a probability ρ with which he observes a signal α_p, at a cost $\kappa(\rho)$, where $\kappa'(\rho) > 0, \kappa''(\rho) > 0, \kappa'(0) = 0, \kappa(0) = 0$, and $\kappa'(1) \geq 1$. I assume that the signal received by the principal is correct with probability 1.[8] The objective of the principal is to maximize ex ante social surplus S.

Actions and Contracts The bureaucrat has two actions: (i) whether to exert effort, e, and (ii) what allocation to give the consumer, a. Based on this allocation the consumer sends a message $m \in \{n, c\}$, where the message n means that no complaint is made and message c implies that a complaint has been made.

The principal has two choices. First, he investigates with probability $\rho(a, m)$, where monitoring depends both on whether the customer complains and on the allocation made by the bureaucrat. He can commit to these probabilities. If the investigation turns up evidence that the agent made a mistake, the principal overturns the agent's decision and allocates the correct one. If the investigation concurs with the agent's findings, or there is no evidence obtained by the principal, the decision is left unchanged from that suggested by the agent. Second, he chooses a wage contract for the bureaucrat.

This paper concerns the role of investigations on bureaucrat behavior. Two issues naturally arise. First, if bureaucrats wish to avoid investigation, what effect does this have on their behavior and practices? Second, why do bureaucrats wish to avoid investigation? In this section of the paper, I deal only with the first of these questions, by assuming an *ad hoc* reward function where the bureaucrat wishes to avoid investigations. Section 5 provides a fully developed career concerns model to justify this: this model is relegated to a later part of the paper for ease of exposition. Accordingly, in this section I assume that the principal chooses the bureaucrat's wage, w, which consists of a salary w_0 and a penalty if an investigation occurs and she made the wrong allocation. In particular, I consider contracts of the form

$$w = w_0 - \Delta I, \tag{2}$$

[8] This assumption is used to rule out "nuisance complaints", where a customer complains even when he knows that the bureaucrat made the right decision in the hope that the principal will come to the wrong conclusion and overturn the bureaucrat's decision in the consumer's interest. I am largely interested here in cases where complaints are informative of bureaucratic error and so ignore this possibility by assuming that the principal never makes errors after an investigation.

where

$$I = \begin{cases} 1 & \text{if the principal observes } \alpha \neq a, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The salary w_0 is chosen so as to satisfy the worker's participation constraint, and is of little importance here, so it is largely ignored in what follows. No contracts can be offered to the consumer.

The timing of the game is as follows. First, nature assigns α to the consumer and the principal and agent sign a contract which specifies both wages (w_0, Δ) and the monitoring propensities $\rho(a, m)$. Second, the agent exerts effort. Third, the customer and the agent privately observe their signals. Next, the agent proposes an allocation a. Following this, the customer send a message m, i.e., he complains or not. The principal then monitors with probability $\rho(a, m)$ as specified in the contract. If he observes α, he allocates $A = \alpha$ and pays the agent according to the contract above. Otherwise, the agent receives w_0 and the agent's recommendation is implemented.

A Role For Complaints I am interested in cases where complaints are informative of bureaucrat error and, hence, increase the likelihood of oversight. This places a bound on the cost of complaint relative to its benefits.[9] I restrict attention to those cases where there is a cost of complaint which is such that she (at least sometimes) complains if the bureaucrat makes an error but not otherwise. To do this, I assume that the consumer has an arbitrarily small cost of complaining: this guarantees a Bayesian Nash equilibrium where complaints are informative. See Prendergast (2001) for complications that arise when the consumer is imperfectly informed and costs of complaint are non-trivial.

Formally, I characterize the Bayesian Nash equilibria of the model when complaints are informative of bureaucrat error. The objective of the principal is to choose $\rho(a, m)$, Δ, and w_0 to maximize $ES(A; \alpha)$ subject to (i) the effort incentives of the bureaucrat, (ii) the truth-telling incentives of the bureaucrat, (iii) her reservation utility, and (iv) the incentives of the consumer to complain. Each of these incentives is described in turn.

1: The Incentive to Complain The efficiency of monitoring depends on the ability of consumers to credibly alert the principal that a mistake has been made; this is what focuses investigations.

[9]Specifically, if the cost of complaint is very large, the consumer never complains and complaints have no allocative role. Similarly, if the cost of complaint is very small relative to the possible benefits, the consumer always complains if denied the asset: once again, complaints mean nothing since they are not indicative of bureaucratic error.

It is always the case that the principal chooses $\rho(a,c) \geq \rho(a,n)$.[10] First note that as the principal observes α upon investigation, consumers never complain when the bureaucrat makes the correct decision. The only issue is whether they complain when an error has been made. Consider the outcome when a bureaucrat recommends allocation j but the consumer knows that allocation i is socially optimal. If the agent complains, this increases the likelihood of i being ultimately implemented. The agent will complain if the bureaucrat errs by allocating j if

$$V(i,i) \geq V(j,i), \tag{4}$$

where the small cost of complaint is ignored.[11] Let $\gamma(a)$ be the probability that a consumer complains if the bureaucrat incorrectly recommends allocation a. If $\gamma(a) = 1$ for all a, I call this the fully informative case.

2: Effort Incentives Consider the incentives of the bureaucrat to collect better information. If she exerts effort, she is incorrect with probability $1-q(1) < 1-q(0)$. The consumer then complains with probability $\gamma(a)$ and the principal monitors with probability $\rho(a,m)$. The expected utility of the bureaucrat is then given by

$$w_0 - \frac{1-q(e)}{2}[\rho(1,c)\gamma(1) + \rho(1,n)(1-\gamma(1)) + \rho(0,c)\gamma(0) + \rho(0,n)(1-\gamma(0))]\Delta - de. \tag{5}$$

The agent's effort incentive constraint is then given by

$$\frac{q(1)-q(0)}{2}[\rho(1,c)\gamma(1) + \rho(1,n)(1-\gamma(1)) + \rho(0,c)\gamma(0) + \rho(0,n)(1-\gamma(0))]\Delta \geq d. \tag{6}$$

3: Truth-Telling The central problem that the paper addresses is the difficulty in inducing bureaucrats to reveal harmful information truthfully. I distinguish between two cases: (i) when an error always leads to a complaint and (ii) when it sometimes leads to complaint.

[10] Consider any equilibrium where $\rho(a,c) < \rho(a,n)$. First, when the agent is correct, the consumer never complains as there is a cost to complaint. If the agent is incorrect, he also does not complain as this increases the likelihood of being investigated. As a result, complaints reveal no information, and are dominated by choosing $\rho(a,c) \geq \rho(a,n)$.

[11] To see this, note that a complaint is only relevant if the principal monitors and corrects the decision to $A = i$. This occurs with probability $\rho(a,c)$ and changes utility by $V(i,i) - V(i,j)$. If no complaint is made, the monitoring probability is $\rho(a,n)$. Thus the consumer can credibly reveal information on bureaucratic error only if $(\rho(a,c) - \rho(a,n))V(i,i) \geq (\rho(a,c) - \rho(a,n))V(i,j)$ for all i and j, i.e., if (4) holds and $\rho(a,c) \geq \rho(a,n)$.

2.1 Fully Informative Consumers

I begin by considering the case where consumers always complain when the bureaucrat makes an error. As an example, consider the incentives of an optician with a myopic patient. The patient cares only about seeing properly, and will complain if he cannot. In this situation, the optician can be effectively monitored by responding to consumer complaints, as the consumer has the "right" objectives. This is meant to be a benchmark for the following sections because in this case, truth-telling is not a binding problem and the first best outcome is possible. Resolving the truth-telling problem when this is not true is the concern of Section 3.

First consider the case where $V(i,i) \geq V(j,i)$ for all i and j. In this case, $\gamma(a) = 1$ for all a. If this condition holds, truth-telling is not a binding constraint when the contract is otherwise optimal.[12] In the absence of a binding truth-telling condition, the objective of the principal is to

$$max_{\hat{\rho}(a,m),\Delta} q(e) + (1 - q(e))[\hat{\rho}(a,c)) - \kappa(\hat{\rho}(a,c))] - q(e)\kappa(\hat{\rho}(a,n)) - de \tag{7}$$

subject to $(q(1) - q(0))[\rho(1,c) + \rho(0,c)]\Delta \geq d$. Let the probabilility of an incorrect allocation when the action is a and the message is m be given by $z(a,m)$, and let $\rho^*(a,m)$ be the *ex post* optimal monitoring intensity. This is given by

$$\rho^*(a,m) = \kappa^{'-1}(z(.)) \tag{8}$$

When a consumer makes a complaint, there is a conditional probability of 1 that a mistake was made so that the principal sets $\rho^*(a,c) = \kappa^{'-1}(1)$ for all a. When no complaint occurs, the likelihood of an error is 0 and so monitoring occurs with probability $\rho^*(a,n) = 0$ for all a. With these monitoring propensities, simple manipulation of (6) implies that the agent can be induced to exert effort by choosing $\Delta \geq \Delta^*$, where $\Delta^* = \frac{d}{\kappa^{'-1}(1)(q(1)-q(0))}$. Thus, when the consumer complains if and only if the bureaucrat errs, there is no need to distort monitoring to induce effort exertion. This is summarized in Proposition 1.

Proposition 1 *Assume that $V(i,i) \geq V(j,i)$ for all i and j. The optimal contract has the following features: (i) the agent exerts effort, (ii) monitoring propensities are $\rho^*(a,m)$, and (iii) $\Delta \geq \Delta^*$.*

This is the best case scenario with bureaucrats, and arises for two reasons. First, consumers are

[12] Formally, if the consumer complains if and only if the bureaucrat errs, the truth-telling condition of the bureaucrat is that she reports the truth $\alpha = a$ rather than a' only if $(1-q(1))\rho(a,c)+q(1)\rho(a,n) \geq q(1)\rho(a',c)+(1-q(1))\rho(a',n)$ which will always be satisfied below for $q(1) > \frac{1}{2}$ at the ex post optimal level of intervention. In this case, the truth-telling condition has no effect as the bureaucrat is more likely to have a complaint made if she lies.

willing to reveal all relevant information by pointing out errors when the bureaucrat inefficiently chooses $a = 1$ and $a = 0$. Second, the bureaucrat is willing to report truthfully at the ex post efficient levels of oversight. One or both of these is violated below. As such, this provides a benchmark for the following sections.

3 Bureaucratic Constraints

The key step in the efficiency result above is that the consumer must be willing to report bureaucratic error in all states. In most bureaucracies, this is unlikely.[13] To cut down on notation, I consider a simple case where this is so. Specifically, I assume that the preferences of the consumer are independent of the true state where $V(1,1) = V(1,0) = v$ and $V(0,0) = V(0,1) = 0$. The agent simply prefers allocation 1 by v to allocation 0.[14] Again, I consider each incentive constraint in turn.

The Incentive to Complain Now the consumer no longer complains when an error has been made. Instead, the consumer complains only if an error has been made *and* he has been denied the good $a = 0$.[15] Thus, $\gamma(1) = 0$.

Effort Incentives To induce the agent to exert effort, the pricipal must choose Δ in a slightly different manner to Section 2, as the equilibrium monitoring propensities are somewhat different because the consumer no longer points out errors where he was incorrectly allocated the asset. Straightforward calculations show the agent can be induced to exert effort if $\Delta \geq \Delta^{**}$, where[16]

$$\Delta^{**} = \frac{d}{(q(1) - q(0)) \left[\frac{\rho(1,n)+\rho(0,c)}{2}\right]}. \tag{9}$$

Truth-Telling Consider any case where $\Delta > 0$ (as is necessary for effort exertion). Unlike the previous section, truth-telling is by no means guaranteed. To see this, it should be obvious that

[13] For example, not many suspects can be trusted to turn themselves in if the police err in not arresting them, nor do unqualified welfare recipients often send back the checks. Finally, it is rare for a student who receives an undeservedly higher grade to ask for a lower one.

[14] Thus, the lazy student values an A as much as the hardworking student, or the guilty person values being found not guilty as much as a person who is truly not guilty.

[15] To avoid the uninteresting problem of the out-of-equilibrium action where the customer complains when allocated $a = 1$, assume that the principal always monitors in that case.

[16] If the bureaucrat reports truthfully in equilibrium, consider the agent's incentives when $\alpha_a = 0$. In this case, the consumer complains if $\alpha = 1$. This occurs with probability $1 - q(e)$. Conditional on this, the principal monitors with probability $\rho(0, c)$. With complementary probability, the agent is right and there is no complaint, and monitoring occurs with probability $\rho(1, n)$.

when the bureaucrat believes that $a = 1$ is optimal, she has no reason not to act honestly.[17] The case where $\alpha_a = 0$ is more difficult. Here the agent faces a choice between (i) honestly denying the consumer, knowing that with probability $1 - q(e)$, a complaint will be made, and (ii) giving the consumer his preferred choice, which, though inefficient, at least causes no complaint.

If she harms the consumer, her expected penalty is $(1 - q(e))\rho(0, c)\Delta$, as she is penalized only if investigated and found wrong. On the other hand, if she offers the consumer $a = 1$, she is monitored with probability $\rho(1, n)$. In this case, the only way that she is not penalized conditional on an investigation is if she was *wrong* and inadvertantly gave the customer the correct allocation. Thus the expected penalty for capitulating to the consumer is given by $q(e)\rho(1, n)\Delta$. Therefore, the bureaucrat can be induced to harm the consumer only if $(1 - q(e))\rho(0, c)\Delta \leq q(e)\rho(1, n)\Delta$, or

$$(1 - q(e))\rho(0, c) \leq q(e)\rho(1, n) \tag{10}$$

for any $\Delta > 0$.

This innocuous condition generates all the results that follow. The issue is whether the bureaucrat is made better off by giving the customer what he wants or by telling the truth and denying him benefits. If she lies, the consumer never complains, and the principal is left with the difficult problem of monitoring cases where the customer has not done him the favor of pointing out that a mistake has been made. The principal realises that an allocation of $a = 1$ could be because the consumer truly deserved the asset or because he was given it unfairly. As a result, the returns to monitoring are low, because many legitimate cases are pooled with those where the bureaucrat was dishonest.[18] By contrast, if she tells the truth, her allocation is probably right ($q(e) > \frac{1}{2}$). But if she is wrong, the consumer complains, which is very informative that an error has been made. The tradeoff between these effects generates the willingness of the bureaucrat to harm the consumer.

In any equilibrium where the agent exerts effort, the objective of the principal is now to

$$max_{\hat{\rho}(a,m)} \frac{1 - q(1)}{2}[\rho(0, c) - \kappa(\rho(0, c))] + \frac{1}{2}[(1 - q(1))\rho(1, n) - \kappa(\rho(1, n))] - \frac{q(1)}{2}\kappa(\rho(0, n) \tag{11}$$

subject to (10) and $\Delta \geq \Delta^{**}$. (This program already incorporates the incentives to complain and exert effort.)

Note that (10) is independent of $\Delta > 0$, so that the resolution of the truth-telling problem

[17] This allocation is preferred by the customer, and since $q \geq \frac{1}{2}$, the bureaucrat knows that the likelihood of a complaint is lower if she offers the right allocation.

[18] For example, a police chief would find sampling all people who were not arrested to be a relatively inefficient way of finding out which suspects should have been arrested.

is independent of the effort exertion problem for any positive Δ. Given this, can the bureaucrat be induced to tell the truth when the principal monitors in the *ex post* efficient way? Call this allocation the second best. Second-best monitoring implies that

$$\rho^{**}(1,n) = \kappa'^{-1}(1-q(e)), \rho^{**}(0,c) = \kappa'^{-1}(1), \rho^{**}(0,n) = 0. \tag{12}$$

Note the lower propensity to monitor when the asset is given to the agent $a = 1$ and there is no complaint in (12) than in the fully informative case. This is because the principal now realises that the allocation could have been correct when no complaint is made. Then the second best truth-telling constraint is given by

$$(1-q(1))\kappa'^{-1}(1) \leq q(1)\kappa'^{-1}(1-q). \tag{13}$$

If this condition holds, the bureaucrat can be induced to report honestly at no cost. If so, monitoring propensities are as in (12) and wages are set as in (9). I call these allocations *non-bureaucratic.*

Proposition 2 *Assume that $V(1,1) = V(1,0) = v$ and $V(0,0) = V(0,1) = 0$ but that (13) holds. Then monitoring propensities are set at their ex post optimal levels $\rho^{**}(a,m)$.*

Note that this entails a distortion relative to the case in Section 2 where the consumer always reports bureaucratic error. Here consumers will not do so if mistakenly given benefits, which implies that the principal now monitors cases where $a = 1$ not knowing whether a correct or incorrect decision was made. When consumers can be trusted to always report bureaucratic error, these cases can be distinguished, whereas here the principal monitors in all cases where $a = 1$ rather than only those where a mistake was made. This yields the first cost of bureaucracies.

My main interest here is not in this cost of bureaucracy, but in the harmful incentives that the truth-telling constraint causes. Note that the second best truth-telling constraint holds if the need to monitor when $a = 1$ is high enough. But there is no reason why it should be. It could be that the (ex post) need to monitor when the customer is given the good is low. In that case, the temptation to accede to the customer is too great, and the second best allocation cannot be obtained. To see this, one example that will be used for comparative statics below is where the costs of monitoring are quadratic: $\kappa(\rho) = \frac{\rho^2}{2}$. In that case, (13) simplifies to $1 \leq q(1)$, which is always violated. Thus, in the quadratic costs case, the bureaucrat can never be induced to report honestly if the bureaucracy monitors (ex post) efficiently.

3.1 The Optimal Bureaucracy

The temptation to accede to consumer demands is characterized by (10). Ironically, the resolution of the bureaucratic agency problem requires that they not feel too threatened by complaints which highlight their mistakes! Remember that to truthfully deny the customer, the bureaucrat must not worry too much about a complaint, yet worry sufficiently about oversight if he gives the good to the consumer. There are two immediate ways to do this: (i) to ignore some complaints, and (ii) to monitor more in the absence of complaints. I illustrate below that where (13) is violated, optimal bureaucracies do both. Proposition 3 describes the optimal contract.

Proposition 3 *Assume that* $V(1,1) = V(1,0) = v$ *and* $V(0,0) = V(0,1) = 0$ *but that (13) is violated. The optimal contract* $(\hat{\rho}(a,m), \hat{w}_0, \hat{\Delta})$ *has the following features:*

1. *If the agent exerts effort, (i)* $\hat{\rho}(0,c) < \rho^*(0,c)$*, (ii)* $\hat{\rho}(1,n) > \rho^*(1,n)$*, and (iii)* $\hat{\rho}(0,n) = \rho^*(0,n)$*. Furthermore,* $\hat{\rho}(0,c)(\hat{\rho}(1,n))$ *is increasing (decreasing) in* q*.*

2. *If the agent exerts no effort, (i)* $\hat{\rho}(0,c) = \rho^*(0,c)$*, (ii)* $\hat{\rho}(1,n) = \rho^*(1,n)$*, and (iii)* $\hat{\rho}(0,n) = \rho^*(0,n)$*. Furthermore,* $\hat{\rho}(0,c)(\hat{\rho}(1,n))$ *is independent of (decreasing in)* q*.*

Proposition 3 illustrates the difficulty of providing incentives in bureaucracies. When the second best is not possible because the bureaucrat accedes to the consumer, the principal is left with two choices. First, he can simply give up on providing incentives. The cost of this is obvious, but at least the bureaucrat has no incentive to lie and so the principal can monitor with the optimal monitoring probabilities, though here with poorly informed bureaucrats. The second strategy is to induce effort exertion, but at the cost of distorting monitoring propensities so that (i) truth-telling becomes less worrisome for the bureaucrat, and (ii) capitulating becomes more costly, even at the expense of the allocative inefficiencies that this entails. Given the convex nature of the problem, this is done by choosing both $\hat{\rho}(0,c) < \rho^*(0,c)$ and $\hat{\rho}(1,n) > \rho^*(1,n)$.[19]

Dealing With Incompetent Bureaucrats Note also from Proposition 3 that the principal responds to complaints about incompetent bureaucrats ($q(1)$ low) less aggressively than to complaints about their more competent counterparts. This runs counter to the economic logic that allocations which are more likely to be wrong should induce greater oversight. But if the agent is induced to exert effort, $\hat{\rho}(0,c)$ is increasing in $q(1)$. When effort incentives are not relevant, this is

[19]There is considerable evidence of a reluctance by police officers to investigate complaints, and institutions like the INS and IRS are hardly renowned for their accessability when errors have been made. See New York Times, 2000, Cannon, 2000, and Christian Science Monitor, 2000, for details.

not the case.[20] To understand the reason for this counterintuitive result, note that where (13) is violated, the truth-telling condition binds in equilibrium so that $(1-q(1))\hat{\rho}(0,c) = q(1)\hat{\rho}(1,n)$ or

$$\frac{q(1)}{1-q(1)} = \frac{\hat{\rho}(0,c)}{\hat{\rho}(1,n)}. \tag{14}$$

Thus the lower is $q(1)$, the lower must be $\frac{\rho(0,c)}{\rho(1,n)}$, which is satisfied by both decreasing $\rho(0,c)$ and increasing $\rho(1,n)$. A more intuitive explanation is as follows. Bureaucrats who are relatively sure of the optimal allocation do not worry much about the prospect of being overturned after an investigation, and so monitoring propensities need to change little from their efficient levels to induce truth-telling. By contrast, when bureaucrats have little idea of the optimal allocation, they are more worried about an investigation and so are more likely to accede to the consumer. To counter this, the principal ignores more valid complaints leveled at an incompetent bureaucrat, though does monitor her more when no complaint is made.[21]

3.2 Giving Up on Incentives

So far I have characterized the equilibria which occur conditional on an effort level. This illustrates that the bureaucratic constraints and practices outlined in the paper are the result of an attempt to induce the agent to exert effort. But I have thus far said nothing about how the principal chooses whether to induce effort exertion. This tradeoff offers little unintuitive. The principal simply trades off the benefits of greater effort ($q(1)-q(0)$ relative to d) against the costs of distorted monitoring ($\hat{\rho}(a,m) - \rho^{**}(a,m)$) when choosing whether to incur the costs of inducing effort exertion. There is little surprising in these results, where lower returns to effort or greater costs of distorted monitoring and decision-making result in the principal being less likely to induce effort exertion. As a result, I do not include formal results here.

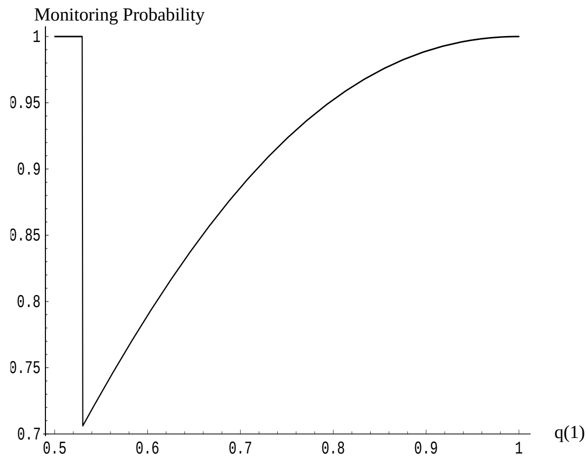
Instead, I summarize the results of this section with an example to illustrate the importance of the bureaucrat's effort and its effect on bureaucratic intervention. Consider a case where the costs of monitoring are quadratic, $\kappa(\rho) = \frac{\rho^2}{2}$, the cost of agent effort is negligible (d close to zero), and in the absence of agent effort, the bureaucrat is uninformed about the true state ($q(0) = \frac{1}{2}$). Without the truth-telling problem, this would imply that the bureaucrat is induced to exert effort. When costs are quadratic, remember that the second best is never possible as 13 is always violated. First consider the equilibrium propensity to monitor after a complaint. The ex post optimal probability of

[20] In the proposition, $\hat{\rho}(0,c)$ is independent of $q(1)$ only because the consumer is perfectly informed. If the consumer were not perfectly informed, the ex post optimal monitoring propensity would be decreasing in $q(1)$.

[21] Note also that these distortions are specific to the problems of inducing effort exertion: in the case where there is no need to induce effort, the principal simply monitors in the non-bureaucratic way.

monitoring is $1(=\kappa'^{-1}(1))$. Figure 1 plots the equilibrium propensity to monitor after a complaint as a function of the quality of the bureaucrat's information.[22]

Figure 1: Monitoring Propensity After a Complaint



First note that if the quality of the bureaucrat's information ($q(1)$) is low, monitoring is set at its ex post efficient level of 1. This is because the principal does not induce effort exertion if its returns are low, and so from Proposition 3 investigates efficiently. Thus even though effort is (almost) costless, it is not exerted. In this example, $q(1)$ must exceed 0.53 before effort is induced. Above this level, the principal induces effort exertion, but at the cost of distorted monitoring. In this example, monitoring after a complaint is initially reduced by 30%, but this converges back towards the efficient level as the agent becomes more certain. Figure 2 provides analogous information on the propensity to monitor when the consumer is allocated the asset, and so there is no complaint.

Figure 2: Monitoring Propensity When the Consumer is Allocated the Asset

Here the ex post efficient monitoring probability is $1-q(1)$, given by the hashed line in Figure 2. However, the truth telling constraint has two effects. First, when the agent is not induced to exert effort ($q(1) < 0.53$), monitoring occurs with probability $1 - q(0) = \frac{1}{2}$ (as the agent is uninformed in the absence of exerting effort). When effort is induced, the propensity to monitor exceeds its ex post optimal level ($\hat{\rho}(1, n) > 1 - q(1)$) for the reasons described above. Note once again that

[22]In this case, the surplus from inducing no effort is $\frac{11}{16}$ and the surplus from inducing effort, at the cost of distorted monitoring, is given by $q(1) + \frac{1-q(1)}{4(1-q(1)+q(1)^2)}$.

monitoring converges to its efficient level as the bureaucrat becomes more certain.

4 Extensions

4.1 The Frequency and Efficiency of Bureaucratic Decision-Making

The model also offers implications for the aggregate frequency and efficiency of oversight. In order to do these comparative statics, I consider the quadratic cost case introduced above, where the truth-telling constraint (13) must be violated. In that case, it is simple to show that if the agent is induced to exert effort, the allocations are given by

$$\hat{\rho}(1,n) = \frac{1-q(1)}{1-q(1)+q(1)^2} \geq 1-q(1) = \rho^{**}(1,n) \tag{15}$$

and

$$\hat{\rho}(0,c) = \frac{q(1)}{1-q(1)+q(1)^2} \leq 1 = \rho^{**}(0,c). \tag{16}$$

The ex ante probability of monitoring is then simply computed to be $\frac{1}{2}\hat{\rho}(1,n) + \frac{1-q(1)}{2}\hat{\rho}(0,c) = \frac{1-q(1)^2}{2(1-q(1)+q(1)^2)}$. This is greater than the level which occurs in the absence of needing to distort monitoring, which is given by $1-q(1)$. Thus, bureaucratic allocations involve more monitoring.

Yet monitoring is ill-targeted for the reasons described above, so a natural question is how this pattern of increased if ill-focused monitoring affects the number of mistakes that are corrected. Mistakes occur with probability $1-q(1)$ so the fraction of mistakes that would be corrected in the absence of distorted monitoring is

$$\frac{1-q(1)}{2}(1) + \frac{1-q(1)}{2}(1-q(1)) = \frac{(1-q(1))(2-q(1))}{2}.$$

When these agency problems imply that monitoring is given in (15) and (16), the analogous correction rate is

$$\frac{1-q(1)}{2}\left(\frac{q(1)}{1-q(1)+q(1)^2}\right) + \frac{1-q(1)}{2}\left(\frac{1-q(1)}{1-q(1)+q(1)^2}\right) = \frac{1-q(1)}{2(1-q(1)+q(1)^2)}.$$

For $q(1) > \frac{1}{2}$, $\frac{(1-q(1))(2-q(1))}{2}$ exceeds $\frac{1-q(1)}{2(1-q(1)+q(1)^2)}$, so that the fraction of mistakes corrected is lower in the bureaucracy. Thus, although bureuacracies monitor more, they correct fewer mistakes.

4.2 Consumer-Biased Oversight

One implication of these bureaucratic constraints is that intervention occurs "too little" when the consumer is denied the asset and "too much" when the consumer is given the asset. The ex ante probability of being given the asset is

$$\frac{q(1) + (1 - q(1))\hat{\rho}(1, n) + (1 - q(1))(1 - \hat{\rho}(0, c))}{2}.$$

Since $\hat{\rho}(0, c) < \rho^{**}(0, c)$ and $\hat{\rho}(1, n) > \rho^{**}(1, n)$, this implies that organizations which monitor in bureaucratic ways deny the asset to the customer more than when monitoring is second best optimal. Thus, oversight of bureaucrats is biased against consumers, in that they are more likely to be denied the asset than in the second best allocation.

4.3 Delay

Bureaucracies are often accused of being too slow, with long apparently unnecessary delays. I argue that such delay may be an efficient way of inducing bureaucrats to reveal information truthfully. When choosing the timing of decisions, individuals usually trade off the cost of delayed receipt of goods against the fact that they may accumulate more information by waiting. This effect is as true for bureaucracies as for other firms. But one feature that distinguishes bureaucracies is the importance of agent certainty for revealing information truthfully. Remember that the bureaucrat can be induced to reveal information truthfully only if she does not worry much about the possibility of a complaint. One way to encourage this is to make bureaucrats more sure of themselves before they make decisions. By delaying decision-making further, the bureaucrat collects more information so that she is (i) less likely to be found wrong if she tells the truth, and (ii) more likely to be penalized if she lies. This additional constraint of inducing truth-telling by bureaucrats results in more delayed decision-making in bureaucracies.

To see this, assume that the principal can now specify a length of time t that bureaucrats must take to make their decisions. This can be contracted on. The value of time is that it makes for better decisions: let the probability that the agent is correct now be given by $q(e, t)$ where $q_2(e, t) > 0, q_2(e, 0) = \infty, q(1, t) - q(0, t) > 0, q_{22}(e, t) < 0$, and $q_2(1, t) - q_2(0, t) \leq 0$. The cost of delay is that the social surplus is discounted, where the interest rate is given by r.

To begin, define the technologically efficient delay in decision-making. This is the delay which would arise if there were no truth-telling concerns. Let the social surplus at the time of delivery be defined by Γ:

$$\Gamma = \frac{q(e,t)}{2} + \frac{1-q(e,t)}{2}\left(\rho(0,c) - \kappa(\rho(0,c)\right) + \frac{1}{2}\left(\rho(1,n)(1-q(e,t)) - \kappa(\rho(1,n))\right). \tag{17}$$

Then efficient delay $t^*(e)$ is given by

$$r\Gamma = \frac{d\Omega}{dq} q_2(e, t^*(e)), \tag{18}$$

where diminishing returns in $\Omega(q)$ and $q_2(e, t^*(e))$ guarantees uniqueness. The delay $t^*(e)$ gives the optimal time to make decisions in the absence of the bureaucrat's incentive constraint. But the analogous truth-telling condition to (13) is now given by

$$(1 - q(1, t^*(1)))\kappa'^{-1}(1) \leq q(1, t^*(1)\kappa'^{-1}(1 - q(1, t^*(1))). \tag{19}$$

When (19) is satisfied, there is no need to distort the timing of decisions from its optimal level. But when this constraint is violated, the bureaucracy changes the speed with which it makes decisions, as illustrated below.

Proposition 4 *Assume that $V(1,1) = V(1,0) = v$ and $V(0,0) = V(0,1) = 0$ and that (19) is violated. The optimal contract $(\hat{t}, \hat{\rho}(a,m), \hat{w}_0, \hat{\Delta})$ has the following features:*

1. *If the agent exerts effort, (i) $\hat{t}(1) > t^*(1)$, (ii) $\hat{\rho}(0,c) < \rho^{**}(0,c)$, (iii) $\hat{\rho}(1,n) > \rho^{**}(1,n)$ and (iv) $\hat{\rho}(0,n) = \rho^{**}(0,n)$.*

2. *If the agent exerts no effort, (i) $\hat{t}(0) = t^*(0)$, (ii) $\hat{\rho}(0,c) = \rho^{**}(0,c)$, (iii) $\hat{\rho}(1,n) = \rho^{**}(1,n)$ and (iv) $\hat{\rho}(0,n) = \rho^{**}(0,n) = 0$.*

Thus, when the principal induces effort exertion from the bureaucrat, there are what appear to be excessive delays. This arises because the principal is faced with a number of costly instruments with which to induce the bureaucrat to honestly report his information. Both monitoring propensities and delay are used to relax the truth-telling constraint so that bureaucrats are made less worried about complaints by delaying decision-making beyond its technologically efficient level.

5 Why Avoid Investigations?

The results of the previous sections are predicated on the assumption that investigations harm bureaucrats; this generates the need to distort monitoring propensities. In this section I provide a career concerns interpretation of the bureaucrat's desire to avoid the limelight. It arises because the agent's performance becomes more visible when an investigation occurs. I show that such limited observability generates the same incentives as in (10), and that similar bureaucratic responses ensue.[23]

To model this, assume that bureaucrats differ in their ability, where ability is unknown to all parties. Bureaucrats are of two types; those who are always right, and those who are always wrong. The fraction of agents who are always right is given by $q(e)$.[24] The bureaucrat's wage depends on perceptions of her ability held by the "labor market", and is given by $E[q(e)|\Omega]\Delta$, where Ω is the information available to the labor market[25]. The bureaucrat's objective is to maximize $E[q(e)|\Omega]\Delta - de$.

A Neutral Benchmark What is important for understanding the effect of investigations on bureaucrat behavior is Ω, the information available to the labor market. I begin by considering a benchmark which offers an efficient allocation; where the market observes all the information available to the principal, namely, the recommendation made by the agent, whether a complaint occurs, whether an investigation is carried out, and the outcome of the investigation.

In this case, she can always be induced to reveal information truthfully without distorting monitoring propensities. The reason for this is simple: investigations are simply a way of revealing the ability of the bureaucrat, and the since the (risk neutral) bureaucrat is indifferent to a mean-preserving spread of perceptions of her ability, she is indifferent to whether an investigation occurs. Consider an equilibrium of the game where the agent is assumed to exert effort and report truthfully, and monitoring propensities are set at the ex post optimal level. Then if she reveals information truthfully, her expected utility is $q(1)\Delta$. If she deviates and offers the consumer his preferred choice, she earns expected utility of $[q(1) - (2q(1) - 1)\rho^*(1, n)]\Delta < q(1)\Delta$.[26] Thus, with this model of career concerns, the distortions of bureaucracy disappear.

This result should not be surprising, as the career concerns model has no reason to make

[23] One, more primitive, reason why bureaucrats have an incentive to avoid the revelation of information about their ability is risk aversion. It is well known (Holmstrom and Ricart i Costa, 1986) that risk averse agents avoid activities that reveal information about their talents when returns are linear in perceptions.

[24] Thus, the return to effort is to increase the fraction of agents who make correct decisions.

[25] The "labor market" really refers to any external constituencies that can affect the welfare of the bureaucrat. For instance, a police officer accused of using excessive force may face costs in the form of jail time, which is generated by pressure from the population rather than another potential employer.

[26] If an investigation occurs and the agent is right, she will be paid $\overline{w} = \Delta$. If an investigation occurs and the agent is wrong, she will be paid 0. If the bureaucrat recommends $a = 1$ and no investigation occurs, there is no change in the perception of the bureaucrat, so she is paid $q(e)\Delta$. Finally, if she offers $a = 0$ and there is no complaint, this implies that she was right and so earns $\overline{w} = \Delta$, while if a complaint is made, she makes 0. Applying these wages to the appropriate likelihood of their occurrance yields these expressions.

bureaucrats avoid investigation. I now show that with a slight change in observability assumptions, this is no longer true.

Limited Observability of Bureaucratic Success The model above assumes that the successes of bureaucrats are as visible as their failures. But this seems unlikely in many cases, because the attention of external constituencies will be piqued by an investigation.[27] In this section, I consider the implication of this limited observability, by assuming that the labor market observes the performance of the bureaucrat only when an investigation occurs. If an investigation occurs, the labor market observes the same information as the principal. Otherwise, no information arises.

In this situation, the market must make an inference about the bureacrat's ability when no investigation occurs. This represents a good signal on the agent's ability if the principal responds to complaints by consumers. As a result, it is not clear why limited observability matters; can't the market observe that the absence of investigations reveals good information on the bureaucrat which balances the effects that arise when investigations actually occur? I show here that although the market does update in this way, it remains the case that bureaucrats wish to avoid investigations.[28]

There are now three relevant wages that the agent can earn. First, if an investigation occurs and she is vindicated, she earns $\overline{w} = \Delta$, while if the investigation occurs and she is wrong, she earns 0. However, if no investigation occurs, she earns w_0, which is given by

$$w_0 = \Delta[\frac{q + q(1-\rho(1,n))}{q + (1-\rho(1,n)) + (1-q)(1-\hat{\rho}(0,c)}], \tag{20}$$

where the term in brackets is the probability of being correct conditional on no investigation.

Then consider whether the the bureaucrat denies the asset to the consumer when she observes $\alpha_a = 0$. If she denies the asset to the consumer, her expected wage is given by $w_0[q + (1-q)(1-\hat{\rho}(0,c)]$. If she gives the asset to the consumer, her expected wage is given by $w_0[1-\hat{\rho}(1,n)] + (1-q)\hat{\rho}(1,n)\Delta$. The truth-telling constraint is then given by

$$w_0[q + (1-q)(1-\hat{\rho}(0,c)] \geq w_0[1-\hat{\rho}(1,n)] + (1-q)\hat{\rho}(1,n). \tag{21}$$

[27]For instance, consider an INS official who admits or refuses entry to the United States. Recent media reports have given considerable visibility to questionable practices where certain minorities are treated incorrectly (New York Times, 2000). These instances are certainly more visible than the cases where petitioners were treated well. Similarly, the performance of the Los Angeles Police Department who used excessive force was more visible than the myriad of cases where arrests were appropriately handled.

[28]See Milbourn, Shockley, and Thakor, 2001, for other work on how limited observability of outcomes affects career concerns problems.

After some manipulation, (21) simplifies to

$$2q \geq 2w_0 - \hat{\rho}(1,n)[2w_0 - 1]. \quad (22)$$

As with (10), this truth-telling condition need not hold at the efficient level of oversight and also exhibits the same features as those in (10). Specifically, the comparative statics of this truth-telling condition are identical to those of (10), in that the truth-telling condition is relaxed by either (i) reducing $\hat{\rho}(0,c)$ below its ex post optimal level, or (ii) increasing $\hat{\rho}(1,n)$ above its optimal level. Hence the insights of the previous section also hold in this career concerns setting.[29]

As an example, consider the quadratic costs case considered above. At the ex post optimal levels of monitoring $\rho^*(a,m)$, the truth-telling condition becomes

$$2q \geq 1 + q^2, \quad (23)$$

which is always violated for $q < 1$, as in the model above. As a result, the principal must change its oversight propensities from the ex post efficient level to induce truth-telling, as in Section 3.

Unlike the incentives in the previous sections, this model offers the bureaucrat the opportunity to be rewarded for making the right decision in addition to being penalized for wrong decisions. Yet still they wish to avoid investigation. This is because investigations are more visible, so that the bureaucrat worries more about outcomes *conditional* on an investigation. In effect, when no investigation occurs and the bureaucrat made the right decision, she does not get all the benefit from this decision, as it could have been that she was wrong but the principal chose not to investigate. But the bureaucrat suffers all the losses if investigated and found to be wrong. Then because investigation occurs more often when the bureaucrat makes an error, she has ex ante incentives to avoid such investigations.[30]

[29] Changing monitoring propensities is a little more complicated than in the model of Section 3. First, there is the direct effect of changing the $\hat{\rho}(0,c)$ and $\hat{\rho}(1,n)$ terms in (21), where decreasing $\hat{\rho}(0,c)$ and increasing $\hat{\rho}(1,n)$ relaxes the constraints in exactly the same way as in (10), leading to the same bureaucratic responses as above. However, there is now an additional indirect effect where w_0, the "no investigation" wage, depends on the monitoring propensities. Note that (22) is relaxed by reducing w_0: truth telling is easier to satisfy when there is less information on ability from not investigating. It is simple to show that w_0 is increasing in both $\hat{\rho}(0,c)$ and $\hat{\rho}(1,n)$ so that the indirect effects lead to fewer investigations in either case. Thus the indirect effect reinforces the direct effect of reducing $\hat{\rho}(0,c)$ but has a countervailing effect on $\hat{\rho}(1,n)$. However, simple algebraic manipulations show that the direct effect always dominates so that the truth-telling constraint is relaxed by reducing $\hat{\rho}(0,c)$ and increasing $\hat{\rho}(1,n)$ as above.

[30] I have motivated the incentives of the bureaucrat in terms of career concerns, which makes fine-tuning of the contract difficult. But there are some alternative contracting assumptions which can alleviate the bureaucratic practices above. First, the principal could offer a bonus for denying the consumer the asset. This possibility, first proposed in Tirole (1986, 1992), becomes a way of giving incentives to act honestly, as acceding to the consumer's demand now costs the bonus to the bureaucrat. In this model, this can be achieved by offering the bureaucrat a bonus of $max\{0, [(1-q(1))\rho^*(0,c) - q(1)\rho^*(1,n)]\Delta\}$ when she offers $a = 0$. As $[(1-q(1))\rho^*(0,c) - q(1)\rho^*(1,n)]\Delta$ is

6 Why Are Bureaucracies Always Inefficient?

So far, I have described reasons why it is problemmatic to use bureaucrats to allocate benefits. So why use them? I make two claims in this section. First, bureaucrats are used when the goods they allocate are not priced: when they are priced appropriately, consumers should be allowed to choose for themselves. The use of bureaucrats should then be limited to cases where it is not efficient to price benefits *ex post*. I show that there are two familiar cases where this is so: (i) where insurance considerations are important, and (ii) where consumer incentives are important. Second, when assets are not priced, bureaucracies *must* involve the inefficiencies described above. Thus, bureaucrats are used only in cases where they are inefficient. I consider the insurance and incentives cases in turn, in order to justify the reduced form preferences given by $V(i,j)$ in Section 2.

6.1 Insurance

A consumer seeks insurance over his health status. A procedure exists that would improve his health, but it is costly. The return to the procedure depends on the severity of his illness. Assume that with probability $\frac{1}{2}$, the consumer is an appropriate candidate for the procedure (severe) ($\alpha = 1$), which brings a utility return $\overline{v}$. With complementary probability, his case is not severe ($\alpha = 0$) and brings a lower return $\underline{v} < \overline{v}$. The monetary cost of the procedure is k. The consumer is perfectly informed about the severity of his illness, while a bureaucrat can only observe his severity with probability $q(e)$ as above. The asset in this case is whether the procedure is done ($A = 1$) or not ($A = 0$).

The consumer is risk averse and would like insurance over the severity of his case. He is assumed to have an indirect utility function $V(A, \alpha) = U(y) + Av(\alpha)$ where y is his net income and α is his

the expected benefit to the bureaucrat from capitulating to the consumer, this eliminates the bureaucrat's incentive to deny the consumer. Second, one could make the penalties depend on the allocation. This unbundles (at least some component of) the penalty from the decision to investigate. For example, consider a contract of the form: $w = w_0 - \Delta(a)I(a)$ where $I(a) = \begin{cases} 1 & \text{if principal observes } \alpha \neq a, \\ 0 & \text{otherwise.} \end{cases}$. Then the truth-telling condition becomes $(1 - q(1))\hat{\rho}(0,c)\Delta(0) \leq q(1)\hat{\rho}(1,n)\Delta(1)$. If contracts of this form are feasible at no welfare cost, the principal can overcome the bureaucrat's incentives to accede to the consumer's desires with no distortions in monitoring, by using larger penalties if the consumer is given the asset than if denied the asset. Third, the principal could reward the agent particularly well for correct decisions. Finally, the principal could reward the agent for having complaints made against her.

These contracts are outside the spirit of this paper, which instead focuses on how career concerns largely dominate the incentives of bureaucrats. Yet there is obviously a role for them. However, each of these contracts is likely to involve some ex ante marginal costs, so that these instruments will in equilibrium substitute partially for the bureaucratic responses above. Thus, while the quantitative effects of bureaucracy will change, the qualitative effects will remain intact.

health status (severe or not). I assume that $U'(y) > 0, U''(y) < 0$ and that m is the consumer's gross income. In order to make the problem interesting, I assume that the procedure is only efficient when the consumer is severely ill:

$$U(m-k) + \underline{v} < U(m) < U(m-k) + \overline{v}.$$

The principal chooses an insurance contract and an allocation mechanism to maximize ex ante welfare subject to breaking even. The contract chosen by the principal consists of (i) a premium, μ and (ii) a copayment when the procedure is done, ν. I allow the principal to choose between two allocation mechanisms: (i) where the consumer chooses whether to have the procedure done, and (ii) where a bureaucrat makes that decision.

6.1.1 Consumer Choice

First assume that the decision to allocate health care is in the hands of the consumer. The consumer must then be induced to reveal his status correctly. The truth telling constraint for the customer is simply that $V(i,i) \geq V(j,i)$, which requires that

$$U(m-\mu-\nu) + \overline{v} \geq U(m-\mu) \tag{24}$$

and

$$U(m-\mu-\nu) + \underline{v} \leq U(m-\mu). \tag{25}$$

If these two conditions hold, the consumer reports truthfully. The objective of the principal is then to maximize ex ante welfare $\frac{1}{2}U(m-\mu-\nu) + \frac{1}{2}\overline{v} + \frac{1}{2}U(m-\mu)$ subject to (24), (25), and $\mu + \frac{1}{2}\nu \geq k$. Let the optimum be given by (μ^*, ν^*).[31] The inefficiency from this contract compared to perfect insurance is given by

$$\frac{1}{2}U(m-\mu^*-\nu^*) + \frac{1}{2}U(m-\mu^*) - U(m-\frac{k}{2}) > 0, \tag{26}$$

as perfect insurance implies a premium of $\frac{k}{2}$. This is the deadweight loss from allowing the consumer to choose his procedure.

[31] The optimum is characterized by the smallest pair of prices (μ, ν) such that $U(m-\mu^*-\nu^*) + \underline{v} = U(m-\mu^*)$ and $\mu^* + \frac{1}{2}\nu^* = k$. Define $\tilde{\nu}(\mu)$ to satisfy $U(m-\mu-\tilde{\nu}) + \underline{v} = U(m-\mu)$. By risk aversion, this is decreasing in μ. Then the principal simply increases μ until $\mu^* + \frac{1}{2}\tilde{\nu}^*(\mu^*) = k$. This characterizes the optimal contract.

6.1.2 Bureaucrat Choice

Bureaucrats are only used when the consumer cannot be trusted to choose the right allocation, as the consumer is better informed than the bureaucrat. But truthful revelation by the consumer requires that $V(i,i) \geq V(j,i)$, which is the same condition that determines whether the consumer *complains* efficiently. As a result, if the bureaucrat is used, it must be that in equilibrium the consumer only complains when denied the asset in which case there is no reason to offer incomplete insurance to the consumer. This is the benefit of bureaucratic allocation.

Let $\lambda(a,\alpha)$ be the conditional probability that the consumer receives asset a when the severity of the disease is given by α. Remember that this depends on how the principal responds to complaints. Then social welfare from bureaucratic allocation (where μ^{**}, the optimal premium, is chosen to break even) is given by

$$U(m-\mu^{**}) + \frac{1}{2}\lambda(1,1)\overline{v} + \frac{1}{2}\lambda(1,0)\underline{v} \tag{27}$$

What matters for efficiency are the λ's. An efficient allocation mechanism would entail $\lambda(1,1) = 1$ and $\lambda(i,j) = 0$ for all other i and j. As should be clear by now, this does not arise as the bureaucrat makes (both reported and unreported) mistakes and the principal does not always intervene to correct them. Following the results of the previous sections, it should be clear that

$$\lambda(1,1) = q(e) + (1-q(e))\hat{\rho}(0,c), \lambda(0,1) = (1-q(e))(1-\hat{\rho}(0,c)), \lambda(0,0) = q(e) + (1-q(e))\hat{\rho}(1,n)$$

and

$$\lambda(1,0) = (1-q(e))(1-\hat{\rho}(1,n)).$$

The final issue to calculate social welfare from the bureaucratic allocation is the premium. This differs from $\frac{k}{2}$ (the efficient level) for two reasons. First, sometimes the procedure is done when it should not be and vice versa. Second, the principal incurs monitoring costs and the agent incur effort costs which must be paid for. The combination of these events implies that the premium is given by $\mu^{**} = [\frac{1}{2}\lambda(1,1) + \frac{1}{2}\lambda(1,0)]k + \frac{1}{2}(1-q(e))\kappa(\rho(0,c)) + [\frac{1}{2}q(e) + \frac{1}{2}(1-q(e))]\kappa(\rho(1,n)) + de.$[32]

[32]If further pricing of assets is allowed, there is a better allocation than this, where the consumers has an option based on the recommendation of the bureaucrat. Specifically, in the equilibrium above, errors are made, where with probability $\lambda(1,0) = (1-q(e))(1-\rho(1,n))$ an allocative inefficiency arises where the less needy consumer is offered the good. But the principal could simply buy out that consumer's option for the procedure by offering him a payment η_1 for not carrying out the procedure. The optimal value of η_1 is given by $U(m-\mu) + \underline{v} = U(m-\mu+\eta_1)$. In this way, the allocative inefficiency of the procedure can be overcome. Equally, with probability $\lambda(0,1) = (1-q(e))(1-\rho(0,c))$, the consumer is denied the procedure when he needs it, and he can be offered the option to purchase it at a price η_2, where $U(m-\mu) + \underline{v} = U(m-\mu-\eta_2)$. (This is the lowest price at which the less severe case will not purchase.) In this way, the allocative inefficiencies can be overcome. But this also involves incomplete insurance though it eliminates allocative inefficiencies. The incomplete insurance now comes from the fact that some benefit from the bureaucrat's decisions (those incorrectly given the asset) and some lose (those incorrectly

Bureaucratic allocations have two problems compared to the first best allocation. First, the principal incurs monitoring costs and the agent effort costs which are reflected in higher premiums. Second, and of more interest here, procedures are sometimes done that should not be and vice versa (or if more prices are allowed -see footnote 33- there is incomplete insurance to overcome this problem). The distorted allocation occurs both because consumers do not complain when they benefit incorrectly (the distinction between first and second best), and because the principal distorts monitoring to induce the bureaucrat to reveal information truthfully (the bureaucratic constraints of Section 3). These costs must then be compared to the incomplete insurance offered by allowing the consumer to choose. Not surprisingly, if insurance considerations are not so important, the consumer should be allowed to choose. Specifically, bureaucratic allocations are preferred if

$$U(m-\mu^{**})+\frac{1}{2}(\lambda(1,1)-1)\overline{v}+\frac{1}{2}\lambda(1,0)\underline{v}\geq\frac{1}{2}U(m-\mu^{*}-\nu^{*})+\frac{1}{2}U(m-\mu^{*}). \tag{28}$$

Hence the ex ante cost of pricing goods is the determinant of whether bureaucrats should be used.

The Inefficiency of Bureaucracies It should be obvious that the optimal allocation mechanism depends on the ex ante cost of pricing. If the (utility) cost of the co-payment required to ensure truth-telling is not large, the good should be allocated by consumer choice. Inspecting the efficiency of these allocations then paints a picture of organizations which give customers what they want, when they need them (though at some cost). By contrast, once the co-payment cost become sufficiently large, the bureaucrat should allocate the good. But whenever the bureaucrat allocates the good, inefficiencies must arise, as the condition that determines efficient complaints ($V(i,i)\geq V(j,i)$ for all i and j) also determines whether the bureaucrat should be used. Bureaucracies involve inefficiencies if consumers cannot be relied on to always report errors, so bureaucracies must be inefficient when they are used.[33] To phrase this another way, when bureaucracies work

---

turned down). With this pricing mechanism, the expected utility of the agent is given by $\tau[\lambda(1,1)U(m-\mu^{**})+\lambda(1,0)U(m-\mu^{**}-\eta_2)]+\frac{1}{2}[\lambda(0,0)U(m-\mu^{**})+\lambda(0,1)U(m-\mu^{**}+\eta_1)]$, where the premium μ is now given by $\mu^{**}=[\frac{1}{2}k+\frac{1}{2}(1-q(e))\kappa(\rho(0,c))+\frac{1}{2}(1-q(e))\kappa(\rho(1,n))+de-\lambda(1,0)\eta_2-\lambda(0,1)\eta_1]$. Thus, if pricing of these assets is feasible, the allocative inefficiencies of bureaucratic decision-making can be overcome, but at the cost of incomplete insurance.

[33] How inefficient then depends on the truth-telling conditions above. If truth-telling is satisfied at the optimal monitoring levels (the analogous condition to (13)), there is no need to distort monitoring but inefficiencies still arisee because consumers cannot be trusted to always complain when a mistake is made. By contrast, if bureaucrats are not willing to report truthfully at the ex post efficient monitoring levels, further inefficiencies arise. The principal responds in the same way as in Section 3, where he fails to respond to legitimate complaints, monitors in a haphazard fashion, corrects fewer mistakes than he "should", and makes decisions slowly. In the quadratic costs case of Sections 2 and 3, this is the only case that is relevant: whenever the customer cannot be trusted to choose honestly, the bureaucrat always uses the inefficient practices described above.

well, consumer choice works even better. By contrast, when bureaucratic choice works inefficiently, consumer choice works even worse. Thus bureaucratic organizations appear to work less well than those where consumers have more choice, yet here this is no fault of the bureaucracy. Instead, it is simply the nature of the goods that bureaucrats allocate.

6.2 Incentives

Another area where bureaucrats are necessary is that of incentives, such as where a police officer determines whether we should be arrested, a professor decides on the grade of a student, and so on. However, unlike the insurance problem, it is difficult to induce self-selection in most incentive problems. In the context of incentive problems, the asset allocated by a bureaucrat is whether she believes that the customer performed *well* (not guilty, a passing grade, no fault in an auto accident) or *poorly* (guilty, a failing grade, or being at fault in an accident). Think of some relevant incentive situations, and the ability to use consumers to self-select. First, can contracts be designed such that the guilty arrest themselves while the not guility set themselves free? This seems unlikely; if anything, the guilty probably have more reason to fear arrest. Second, could a professor write a contract with students such that the lazy ones choose a D grade while the good ones choose an A? In these examples, I see no reason to expect any single crossing properties to work in the correct direction. As a result, bureaucrats must be used to allocate assets in most incentive situations, resulting in the bureaucratic allocations described above.

7 Conclusion

Bureaucrats certainly get bad press. Disparaging the contributions of penpushers and beancounters has become part of private discourse, media reports, and political campaigns. There is little doubt that bureaucracies have problems for a variety of reasons,[34] but this paper offers a more benign view of the supposed inefficiencies of bureaucrats, where limitations on their ability to allocate resources are largely determined by the nature of the goods that they allocate. This is true even when superiors are armed with the kind of tools that usually guarantee efficient outcomes in other settings.

[34] For example, in many public sector settings, wages cannot be changed by superiors, promotion is by seniority, firing is next to impossible, and objectives for superiors impossibly vague. These insurmountable constraints surely explain much of the inefficiency of public bureaucracies. This paper is not concerned with these problems of public ownership *per se* but rather difficulties that arise from the nature of the jobs that bureaucrats are allocated to do. It is also worth emphasizing that the paper is not meant to address only public sector bureaucracies. The theory offered here considers how agency problems arise when assets are not priced, and are meant to apply as much to the practices of insurance company officials as to police officers.

The central difficulty with bureaucrats is that they work in settings where there are no reliable output measures on which to base pay. Given this problem, superiors rely on customer complaints to focus their attention, as customers are often informed of mistakes which can then be corrected. In a world where consumers pay for assets, there is likely a close association between the objectives of superiors and those of customers, and complaints work well to eliminate agency problems and honest mistakes. But the key problem with bureaucracies is that consumers do not pay for assets, nor should they in some circumstances. This implies that bureaucrats tend to be used in those situations where consumers cannot be fully trusted to reveal errors by bureaucrats, namely in the kind of insurance and incentive problems described above.

But when consumers cannot be trusted to report errors in their favor, bureaucrats have an incentive to accede to their demands even when it is not efficient. The main results of the paper are concerned with overcoming this incentives. How can bureaucrats be given incentives to honestly deny benefits to consumers? I show that the instruments bureaucracies use are unresponsiveness, haphazard monitoring, overall low correction rates, long waiting periods, and allowing particularly incompetent bureaucrats to face little scrutiny after a complaint. The implies a bureaucracy which uses supposedly inefficient rules and says "no" to consumers more than when truth-telling is not a problem. Although this approach surely misses some aspects of incentive provision for bureaucrats (such as those outlined in Dixit, 1997, and Dewatripont et al., 1997), it does at least paint a picture of bureaucratic life that seems more realistic than that offered by the standard agency model.

There is an important asymmetry in the model I presented above. By definition, any incorrect action by a bureaucrat involves losses to some party. But I only allow one party, the consumer, to complain. In some situations, this is unrealistic because other aggrieved parties could also complain about a mistake in the consumers favor. For example, the victim of the crime could complain about the failure of an officer to arrest a suspect. As such, I think this paper is most relevant to those cases where the parties that lose from such mistakes either do not know of them, or are insufficiently organized to voice their grievances. For instance, society loses when someone is inefficiently given unemployment benefits, or receives an unnecessary medical procedure, or is not arrested for selling drugs. Yet the population that is harmed by these actions knows little about what the bureaucrat should have done, and so is unlikely to voice its concerns over this case. The model best reflects these situations, rather than cases where there are clear and well-informed parties on each side of the transaction, where a mistake in either direction will result in a complaint.

Finally, the paper offers a cautionary tale in comparing the efficiency of bureaucracies to other allocation mechanisms. From the perspective of this paper, bureaucracies are used when ex post rents are (efficiently) large enough that consumers cannot be trusted to choose for themselves.

But under the same circumstances, consumers will also not point out bureaucratic error if it is in their favor, which generates the distortions of Section 3. Thus, observed bureaucracies must be inefficient. In other words, bureaucracies work better when consumers complain efficiently (as in Section 2), but they are not necessary in these circumstances. At its simplest, when bureaucracies work well, consumer choice fares even better. As a result, care should be taken in criticizing the performance of bureaucrats: it is true that the police may be less responsive to complaints than a department store, but this is hardly comparing like to like.

Appendix: Proofs of Results

Proof of Proposition 3:

First consider the case where $e = 1$. Conditional on truth-telling, the bureaucracy chooses Δ large enough to induce truth-telling. The key issue becomes the cheapest way to induce the agent to report truthfully, which requires that (10) holds. The objective of the principal is then to

$$max_{\rho(a,m)} \frac{1-q(1)}{2}[\rho(0,c)-\kappa(\rho(0,c))]-\kappa(\rho(0,n))+\frac{1}{2}[(1-q(1))\rho(1,n)-\kappa(\rho(1,n))]. \tag{29}$$

subject to (10). When (10) is violated at the optimal monitoring propensities, the principal chooses monitoring propensities such that (10) binds. By substitution, the objective function is

$$max_{\rho(a,m)} \frac{1-q(1)}{2}[\rho(0,c)-\kappa(\rho(0,c))]-\kappa(\rho(0,n))+\frac{1}{2}[\frac{(1-q(1))^2}{q}\rho(0,c)-\kappa(\frac{(1-q(1))}{q}\rho(0,c))]. \tag{30}$$

Straightforward differentiation yields

$$\hat{\rho}(0,n)) = 0, \tag{31}$$

and

$$q(1)[1-\kappa'(\rho(0,c))]+(1-q(1))-\kappa'(\frac{(1-q(1))}{q}\rho(0,c))] = 0. \tag{32}$$

This equation is optimally satisfied by choosing $\hat{\rho}(0,c)$ such that

$$q(1)\kappa'(\hat{\rho}(0,c))+(1-q(1))\kappa'(\frac{(1-q(1))}{q}\hat{\rho}(0,c))] = q(1)+(1-q(1))^2. \tag{33}$$

It can be easily seen that the optimal allocation implies that $1 > \kappa'(\hat{\rho}(0,c)) > q$. At $1 = \kappa'(\hat{\rho}(0,c))$ (the ex post optimal level), the first order condition is characterized by $\kappa'(\frac{(1-q(1))}{q}\hat{\rho}(0,c))] > (1-q(1))$. Similarly, at $\kappa'(\hat{\rho}(0,c)) > q$, which implies that $\kappa'(\hat{\rho}(1,n)) > 1-q$ (the ex post optimal level), the first order condition is given by $\kappa'(\hat{\rho}(0,c)) < 1$. Since the marginal cost functions are monitonically increasing in monotoring propensities, this implies that $1 > \kappa'(\hat{\rho}(0,c))$ and $\kappa'(\hat{\rho}(1,n)) > 1-q$.

Next note that the truth-telling constraint binds when (13) is violated. This implies that $\frac{q(1)}{1-q(1)} = \frac{\hat{\rho}(0,c)}{\hat{\rho}(1,n)}$. By the convexity of the marginal cost functions, this implies that $\hat{\rho}(0,c)$ is increasing in $q(e)$ and $\hat{\rho}(1,n)$ is decreasing in $q(e)$.

Next consider the case where $e = 0$. Then it is optimal to choose $\Delta = 0$ because the agent has no reason not to report honestly. In that case the firm chooses its monitoring propensities to

$$max_{\rho(a,m)} \frac{1-q(0)}{2}[\rho(0,c)-\kappa(\rho(0,c))]-\kappa(\rho(0,n))+\frac{1}{2}[(1-q(0))\rho(1,n)-\kappa(\rho(1,n))] \tag{34}$$

subject to the salary w_0 meeting the agent's outside opportunities. This implies that monitoring propensities are given by

$$\kappa'(\hat{\rho}(0,c)) = 1, \hat{\rho}(0,n) = 0, \kappa'(\hat{\rho}(1,n)) = 1-q(0),$$

which are the optimal ex post monitoring propensities conditional on no effort exertion. By inspection, it is clear that $\hat{\rho}(0,c)$ is independent of $q(0)$ and $\hat{\rho}(1,n)$ is decreasing in $q(0)$.

Proof of Proposition 4:

First consider the case where $e = 1$. The objective of the principal is now to

$$max_{\rho(a,m),\tau} e^{-rt} \left(q(1,t) + \frac{1-q(1,t)}{2}[\rho(0,c) - \kappa(\rho(0,c))] - \kappa(\rho(0,n)) + \frac{1}{2}[(1-q(1,t))\rho(1,n) - \kappa(\rho(1,n))] \right) \tag{35}$$

subject to (19). When (19) is violated at the first best, the principal chooses monitoring propensities in exactly the same way as in Proposition (3) except for the notational changes of $q(e,\hat{t})$ for $q(e)$ above. The derivation is thus excluded here.

Now consider the optimal delay. Let the shadow price (Lagrangian) on the constraint for (19) be given by λ and let $\Gamma = \left(\frac{1-q(1,t)}{2}[\rho(0,c) - \kappa(\rho(0,c))] - \kappa(\rho(0,n)) + \frac{1}{2}[(1-q(1,t))\rho(1,n) - \kappa(\rho(1,n))] \right)$. Then the optimal choice of timing to make a decision is given by

$$r\Gamma = \frac{dq(1,t)}{dt} \left(1 - \frac{\rho(0,c) - \kappa(\rho(0,c))}{2} - \rho(1,n) + \lambda(\rho(0,c) + \rho(1,n)) \right). \tag{36}$$

When the truth-telling constraint holds at the first best, this collapses to the efficient level of delay:

$$r\Gamma = \frac{dq(1,t^*)}{dt} \left(1 - \frac{\hat{\rho}(0,c) - \kappa(\hat{\rho}(0,c))}{2} - \hat{\rho}(1,n) \right), \tag{37}$$

where

$$\kappa'(\hat{\rho}(0,c)) = 1, \hat{\rho}(0,n) = 0, \kappa'(\hat{\rho}(1,n)) = 1 - q(1,t^*).$$

But when the truth telling constraint binds, the shadow price $\lambda > 0$ and hence $\hat{t} > t^*$ from (36).

As an example, again consider the quadratic cost case where a closed form solution is possible. Here substitution implies that in the absence of the truth telling constraint, surplus is given by

$$\Gamma_1 = e^{-rt^*} \left(q + \frac{(1-q)(2-q)}{4} \right), \tag{38}$$

while when the constraint exists surplus is given by

$$\Gamma_2 = e^{-r\hat{t}} \left(q + \frac{(1-q)}{4} \frac{1+3q-3q^2}{(1+q-q^2)^2} \right). \tag{39}$$

Then differentiation implies that

$$\frac{d\Gamma_2}{dq} - \frac{d\Gamma_1}{dq} = \frac{3-2q}{4} - \frac{1+3q-3q^2}{4((1+q-q^2)^2} - \frac{1-q}{4} \frac{(2q-1)(1-3q+3q^2)}{(1+q-q^2)^3} > 0, \tag{40}$$

for $q > \frac{1}{2}$. Thus, the marginal return to increasing $q(1,t)$ is higher when truth-telling is a constraint than when there is no such constraint, so that delay is longer as q is concave in t.

Next consider the case where $e = 0$. Then it is optimal to choose $\Delta = 0$ as above. In that case, the firm chooses its monitoring propensities in exactly the same way as in Proposition 3 and chooses delay as in (37), except where $q(0,t)$ substitutes for $q(1,t)$.

References

Becker, G., "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 1968, Vol. 76, pp. 169-217.

Cannon, Lou, "One Bad Cop", New York Times, October 2, 2000, New York.

Christian Science Monitor, "Frisking the Feds", August 30, 2000.

Dewatripont, M., I Jewitt, and J. Tirole, "Missions and the Accountability of Government Agencies", mimeo, 1997.

Dixit, Avinash, "Power of Incentives in Private Versus Public Organizations?", *American Economic Review Papers and Proceedings*, 1997, May, 378-83.

Friebel, Guido and Michael Raith, "Abuse of Authority and Hierarchical Communication," mimeo, Chicago GSB, May 2001.

Milbourn, Todd, Richard Shockley, and Anjan Thakor (2001) "Managerial Career Concerns and Investments in Information," forthcoming, *RAND Journal of Economics.*

Holmstrom, Bengt and Joan Ricart I Costa (1986), "Managerial Incentives and Capital Management", *Quarterly Journal of Economics*, 101, 835-60.

Holmstrom, Bengt and Paul Milgrom (1992), "Multi-Task Principal Agent Analyses: Linear Contracts, Asset Ownership and Job Design", *Journal of Law, Economics, and Organization*, 7, 24-52.

New York Times, "Besmirched Deportland Wrestles With the INS" August 31, 2000a, New York.

Prendergast, Canice, "The Provision of Incentives in Firms", *Journal of Economic Literature*, 1999, 37, 7-63.

Prendergast, Canice, "The Role of Consumers in Agency Problems", forthcoming, *Economic Journal.*

Star Tribune, "Congressional Cures for the IRS Worse than Agency's Ills", August 30, 2000.

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

Tirole, Jean, "Collusion and the Theory of Organizations", in *Advances in Economic Theory: Sixth World Congress, Vol. 2.* J. Laffont, ed. Cambridge: Cambridge University Press, 1992.

Wilson, James, *Bureaucracy*, Basic Books, New York, 1989.