The Limits of Bureaucratic Efficiency

Canice Prendergast

University of Chicago and National Bureau of Economic Research

Bureaucracies tend to be used when consumers cannot be trusted to choose outcomes efficiently. But a primary means of bureaucratic oversight is consumer complaints. But this can give bureaucrats an incentive to inefficiently accede to consumer demands to avoid a complaint. I show that when this incentive is important, bureaucracies (efficiently) respond by (i) ignoring legitimate consumer complaints, (ii) monitoring more in situations in which it is not needed, (iii) delaying decision making "too long," and (iv) biasing oversight against consumers. I also show that bureaucracies are used only when consumers cannot be trusted. As a result, observed bureaucracies are always inefficient.

I. Introduction

Bureaucrats pervade economic life. They approve our medical procedures, process our credit card inquiries, 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.

I am grateful to the editor, Kent Daniel, Lars Stole, two anonymous referees, and seminar participants at Harvard University, University of Michigan, Bristol University, Duke University, the Latin American Econometric Society, the Irish Economic Association, and the Royal Economic Society, Durham, for helpful comments. I would also like to thank the National Science Foundation and the University of Chicago for generous support. Any errors are my own.

[*Journal of Political Economy*, 2003, vol. 111, no. 5]
 0022-3808/2003/11105-0007$10.00

There are typically two ways to assign goods to consumers: consumer choice and bureaucratic allocation. I show two related results in this paper. First, when consumers cannot be trusted to allocate goods efficiently, bureaucracies are necessarily inefficient. Second, bureaucrats are used only when consumers cannot be trusted to efficiently allocate goods. The key insight of the paper then is that efficient bureaucracies are never observed because the features that lead to their use also make them inefficient. Put another way, when bureaucracies work well, consumer choice works better; but when bureaucracies work poorly, consumer choice works worse.

It is hard to find much good that is said about bureaucracies, both private and public. The Internal Revenue Service (IRS) and Immigration and Naturalization Service (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 (bean counters, pen pushers, 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 use 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 important, bureaucrats are used only when they exhibit these "inefficiencies."

I define a bureaucrat as someone who has control over an (observed) allocation to a customer; control derives from private information that she holds over its optimal use. I argue that agency problems from bureaucracy arise for two reasons. First, the decisions made by bureaucrats involve ex post rents to consumers. (For example, it should matter to a patient that he be 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.

But consumers continue to play a role through overseeing the performance of bureaucrats. Bureaucrats are rarely offered pay for performance based on easily available outcome measures. Nor are they offered rewards based only on the allocations they propose. For instance,

police officers are not rewarded when they arrest someone, nor are benefit officials rewarded whenever they deny benefits to a consumer. Instead, the primary way of controlling the behavior of bureaucrats is 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, nor should it be. Instead, investigations are targeted to cases in which a mistake is likely to have been made and are typically triggered by particular signals of a mistake or malfeasance. 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).

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 nonbureaucratic organizations. For instance, poor service in a store will often result in a request to "see the manager." I argue here that complaint mechanisms have particular problems in bureaucratic settings. The reason for this is that typically the consumers of bureaucracies have preferences that do not correspond to 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 to be done, and suspects do not want to be arrested.[1]

Problems then arise 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. I am mainly interested here in a second problem with complaint 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 1989, 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

[1] Political scientists have also considered the use of interested parties for oversight, as in McCubbins and Schwartz (1984) and de Figueiredo, Spiller, and Urbiztondo (1999). Also related are Freibel and Raith (2001) and Leaver (2001).

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* ("Besmirched Deportland Wrestles with the INS," 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" ("Congressional Cures for the IRS Worse than Agency's Ills," 2000).

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 shows 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 complaints.

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 in which 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 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 in which 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 III.

Section IV considers some extensions. 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 overscrutinizing cases with no complaint, they intervene too much when the consumer is (sometimes incorrectly) given the asset. A second implication is that when the cost of monitoring is quadratic,

this policy of ignoring some complaints and overmonitoring routine cases results in (i) a higher average probability of monitoring than in nonbureaucratic settings, but (ii) fewer mistakes are corrected. Thus Section IV offers a theory of bureaucracy that 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 for bureaucrats to be 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 this way, they become more likely to reveal their findings truthfully rather than capitulate to the desires of consumers, because the longer they wait, the more sure they become.

Thus far, the results show that problems arise with bureaucracies only when consumer preferences are not aligned with those of the principal. But an alternative is to allow the consumer to choose his consumption rather than give this power to a bureaucrat. In effect, an alternative to consumer *voice* is consumer *choice*. In Section V, I address whether consumers or bureaucrats should choose consumption. I begin by showing that when consumers can be trusted to allocate goods in the efficient way, allowing consumers to choose dominates using bureaucrats. This is not surprising since I assume that consumers know their own preferences better than bureaucrats do. By contrast, when consumers cannot be trusted, bureaucrats should make the allocation. These observations lead to the central result of the paper, namely, that bureaucrats are used only for "hard" agency problems; the feature that makes consumer choice a poor allocation mechanism (misaligned preferences) also renders consumer voice a poor instrument in the principal's desire to oversee the bureaucrat. Put simply, if consumers cannot be trusted to choose the right allocation, they cannot be trusted to voice their concerns efficiently, which reduces bureaucratic efficiency. As a result, bureaucracies are constrained by the very feature that leads to their existence!

This paper is not meant as an apology for all forms of bureaucratic inefficiency. Instead, it points to situations in which there are limits to how much can be attained by a bureaucracy, namely, those cases in which consumers *optimally* have preferences that diverge from a principal's. I conclude in Section VI by addressing situations in which this is likely. Two cases are highlighted: (i) those for which it is not efficient to price benefits and (ii) those in which benefits are difficult to observe.

II. A Model of Bureaucracy

An allocation A must be made to a consumer, where A can take on a value of zero or one. 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 $\alpha = 0$. I assume that each state occurs with equal probability and that A is observed by all parties.

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 a signal α_a, which is correct with probability $q \geq \frac{1}{2}$. The precision of the agent's estimate depends on her unobserved effort decision: she chooses either high effort $e = 1$ or low effort $e = 0$, where high effort has a disutility of d. Let $q(e)$ be the precision, with $q(1) > q(0)$. 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. 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.

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 α at some cost. Specifically, the principal chooses a probability ρ with which he observes α, at a cost $\kappa(\rho)$, where $\kappa'(\rho) > 0$, $\kappa''(\rho) > 0$, $\kappa''' \geq 0$, $\kappa'(0) = 0$, $\kappa(0) = 0$, and $\kappa'(1) \geq 1$.[2] The objective of the principal is to maximize ex ante social surplus S minus any effort and investigation costs.[3]

Actions and contracts.—The bureaucrat has two decisions: (i) whether to exert effort, e, and (ii) what allocation to give the consumer, a. On the basis of this allocation, the consumer sends a message $m \in \{n, c\}$,

[2] I assume that the signal received by the principal is correct with probability one. This assumption is used to rule out "nuisance complaints," in which 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 in which complaints are informative of bureaucratic error, and so I ignore this possibility by assuming that the principal never makes errors after an investigation.

[3] There are monetary transfers as part of the incentive contract. However, these are simply transfers, and since all parties are risk neutral, they are not part of the efficiency calculation.

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 $\rho(a, m)$; monitoring can depend both on whether the customer complains (m) and on the allocation made by the bureaucrat (a). 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 bureaucrats' behavior. Two issues naturally arise. First, if bureaucrats wish to avoid investigations, what effect does this have on their behavior and practices? Second, why do bureaucrats wish to avoid investigation? In this paper, I deal only with the first of these questions by assuming an ad hoc reward function in which the bureaucrat wishes to avoid investigations. (Think of this as an agent who loses her job only if she is found to have erred.) However, in an earlier article (Prendergast 2002*a*), I provide a fully developed career concerns model to show that the results obtained here are more general: I consider this case here for brevity and expositional simplicity. Here 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

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

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 that 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 sends a message m; that is, he complains or not. The principal then monitors with probability $\rho(a, m)$ as specified in the contract. If he observes the truth, 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.—Complaints have to affect investigations for this model to be meaningful, which requires that they are informative of bureaucrat error and, hence, increase the likelihood of oversight. This places a bound on the cost of complaining relative to its benefits. Specifically, if the cost of complaining is very large, the consumer never complains and complaints have no allocative role. Similarly, if the cost of complaining 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. I restrict attention to those cases in which there is a cost of complaining 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 in which complaints are informative.

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 oversight and wages ($\rho(a, m)$, Δ, and w_0) to maximize the expected surplus from the allocation minus effort and investigation costs, 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.

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. It is always the case that a complaint increases the probability of an investigation ($\rho(a, c) \geq \rho(a, n)$).[4] First note that as the principal observes the truth (α) upon investigation, the consumer never complains 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 the allocation is wrong and that i is correct. If the agent complains, this increases the likelihood that the allocation will be changed, so that he complains if

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

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

the small cost of complaining is ignored.[5] 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. She is penalized if she is incorrect and is investigated. She is incorrect with probability $1 - q(e)$. The consumer then complains with probability $\gamma(a)$, and the principal monitors with probability $\rho(a, m)$. Therefore, the expected penalty with allocation a and consumer response $\gamma(a)$ is $[1 - q(e)]\gamma(a)\rho(a, m)\Delta$. Summing these over the states yields that the agent's expected utility depends on effort, complaints, and investigation propensities, and this is given by

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

If the contract requires that the agent exert high effort, then her effort incentive constraint is 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 leads to a complaint only if the consumer is denied benefits.

A. *Fully Informative Consumers*

I begin by considering the case in which the consumer always complains when the bureaucrat makes an error. (This is meant to be a benchmark for the following sections, because I show below that in this case, allowing consumers to choose dominates using bureaucrats.) When $V(i, i) \geq V(j, i)$, truth-telling is not a binding constraint when the contract is

[5] To see this, note that a complaint is relevant only 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 $\rho(a, c) \geq \rho(a, n)$.

otherwise optimal.[6] I assume that in the absence of these agency concerns, high effort ($e = 1$) is optimal so that the objective of the principal is to

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

subject to the incentive constraint when consumers always report errors, which is

$$\frac{q(1) - q(0)}{2}[\rho(1, c) + \rho(0, c)]\Delta \geq d. \quad (8)$$

Let the probability of an incorrect allocation when the action is a and the message is m be $z(a, m)$, and let $\rho^*(a, m)$ be the ex post optimal monitoring intensity. This is given by

$$\rho^*(a, m) = \kappa'^{-1}(z(a, m)). \quad (9)$$

When a consumer makes a complaint, there is a conditional probability of one that a mistake was made so that ex post optimal monitoring is $\rho^*(a, c) = \kappa'^{-1}(1)$ for all a. When no complaint occurs, the likelihood of an error is zero and so no investigation is needed—$\rho^*(a, n) = 0$. With these monitoring propensities, simple manipulation of (8) implies that the agent can be induced to exert effort by choosing $\Delta \geq \Delta^*$, where $\Delta^* = d/\{\kappa'^{-1}(1)[q(1) - q(0)]\}$. Thus, when the consumer complains if and only if the bureaucrat errs, there is no desire 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 willing to reveal all relevant information. Second, the bureaucrat is willing to report truthfully at the ex post efficient levels of oversight. One or both of these are violated below.

III. 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 many situations,

[6] 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 since the bureaucrat is more likely to have a complaint made if she lies.

this is unlikely. To cut down on notation, I consider a simple case in which this is so. Specifically, I assume that the preferences of the consumer are independent of the true state in which $V(1,1) = V(1,0) = v$ and $V(0,0) = V(0,1) = 0$. The agent simply prefers allocation 1 by v to allocation 0. Thus, for example, the lazy student values an A as much as the hard-working student, or the guilty person values being found not guilty as much as a person who is truly not guilty. 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, he complains only if an error has been made *and* he has been denied the good $a = 0$.

Effort incentives.—If the principal chooses to induce the agent to exert effort, he chooses wages in a manner slightly different from that of Section II since the equilibrium monitoring propensities are different. Here, oversight when the agent allocates $a = 1$ is

$$\Delta^{**} = \frac{d}{[q(1) - q(0)]\{[\rho(1, n) + \rho(0, c)]/2\}}. \tag{10}$$

Truth-telling.—In contrast to the previous section, truth-telling is by no means guaranteed: here there is a temptation to simply give the consumer what he wants. Consider any case in which $\Delta > 0$ (as is necessary for effort exertion). It should be obvious that when the agent believes that the consumer should be given the benefit, she has no reason not to act honestly.[7] The case in which she should be denied ($\alpha_a = 0$) is more difficult. Here the agent faces a choice between (i) honestly denying the consumer, knowing that with probability $1 - q(e)$ she is wrong and a complaint will be made, and (ii) giving the consumer his preferred choice, which, though inefficient, at least causes no complaint.

If the agent harms the consumer, her equilibrium expected penalty is $[1 - q(e)]\rho(0, c)\Delta$ since she is penalized only if she is wrong (with probability $1 - q(e)$) and investigated (with conditional probability $\rho(0, c)$). On the other hand, suppose that she gives the consumer what he wants ($a = 1$) even though her information tells her to deny the benefit. Then there is no complaint and she is monitored with probability $\rho(1, n)$. In this case, she is penalized if she is investigated and her information (that the consumer should be denied) was right, so that the expected penalty for capitulating to the consumer is given by

[7] 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.

$q(e)\rho(1, n)\Delta$. Therefore, the bureaucrat can be induced to efficiently harm the consumer only if $[1 - q(e)]\rho(0, c)\Delta \leq q(e)\rho(1, n)\Delta$, or

$$[1 - q(e)]\rho(0, c) \leq q(e)\rho(1, n) \tag{11}$$

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 in which the customer has not done him the favor of pointing out that a mistake has been made. As a result, the returns to monitoring are low because many legitimate benefit approvals are pooled with those in which the bureaucrat was wrong. 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. By contrast, if the bureaucrat 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 and is likely to result in a penalty. The trade-off between these effects generates the willingness of the bureaucrat to harm the consumer.

To see how this affects monitoring, note that the objective of the principal is now to choose oversight to maximize the surplus from the allocation minus investigation and effort costs,

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

$$- \frac{1}{2}\kappa(\rho(1, n)) - \frac{q(e)}{2}\kappa(\rho(0, n)) - de \tag{12}$$

subject to truth-telling—(11)—and effort incentives ($\Delta \geq \Delta^{**}$ if $e = 1$ is chosen).

First ignore the truth-telling problem. Then the principal monitors in the ex post efficient way. Call this allocation the second-best. Second-best monitoring implies that

$$\rho^{**}(1, n) = \kappa'^{-1}(1 - q(e)), \quad \rho^{**}(0, c) = \kappa'^{-1}(1), \quad \rho^{**}(0, n) = 0. \tag{13}$$

Note the less well focused monitoring when the benefit is given to the consumer than in the fully informative case. The reason is that the principal cannot rely on him to point out errors. Substituting these second-best monitoring levels into (11) implies that truth-telling with the high effort level $e = 1$ is satisfied if

$$[1 - q(1)]\kappa'^{-1}(1) \leq q(1)\kappa'^{-1}(1 - q(1)). \tag{14}$$

If this condition holds, the bureaucrat can be induced to report honestly without distorting oversight, and wages are set as in (10). I call these allocations *nonbureaucratic.* Proposition 2 immediately follows.

PROPOSITION 2. Assume that $V(1, 1) = V(1, 0) = v$ and $V(0, 0) = V(0, 1) = 0$ but that (14) holds. Then monitoring propensities are set at their ex post optimal levels $\rho^{**}(a, m)$ and the agent exerts high effort.

Note that although this outcome is *nonbureaucratic,* it entails an inefficiency relative to the case in Section II in which the consumer always reports bureaucratic error. Here the consumer will not do so if mistakenly given benefits, which implies that the principal now monitors cases in which the benefit is given 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 and monitoring better focused. 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 truth-telling constraint in (14) holds only if the need to monitor when the consumer is allocated the benefit is sufficiently high. The reason is that the way to induce the agent to report honestly is the prospect of being "caught" if she incorrectly gives the consumer the benefit. But there is no reason why this should be so, in which 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 one in which the costs of monitoring are quadratic: $\kappa(\rho) = \rho^2/2$. In that case, (14) 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.

A. *The Optimal Bureaucracy*

The temptation to accede to consumer demands is characterized by (11). Ironically, the resolution of this agency problem requires that the bureaucrat not feel too threatened by complaints that highlight her 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 (legitimate) complaints and (ii) to monitor more in the absence of complaints. I illustrate below that when (14) is violated, optimal bureaucracies do both if they provide incentives to exert effort. 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 (14) 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 $q(1)$.
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) $q(0)$.

Proposition 3 (proved in the Appendix) 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 by choosing $\Delta = 0$. 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. 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 both ignoring legitimate complaints and overmonitoring when no complaints are made.[8]

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 ($q(1)$ high). This runs counter to the economic logic that allocations that are more likely to be wrong should induce greater oversight. An intuitive explanation for this result 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 he does monitor her more when no complaint is made.[9]

[8] There is considerable evidence of a reluctance by police officers to investigate complaints, and institutions such as the INS and IRS are hardly renowned for their accessibility when errors have been made. See "Besmirched Deportland" (2000) and Cannon (2000) for details.

[9] More formally, note that where (14) is violated and effort is exerted, the truth-telling condition binds in equilibrium, so that $q(1)/[1 - q(1)] = \hat{\rho}(0, c)/\hat{\rho}(1, n)$. Thus the lower $q(1)$ is, the lower $\rho(0, c)/\rho(1, n)$ must be, which is satisfied by both decreasing $\rho(0, c)$ and increasing $\rho(1, n)$. Note that these distortions are specific to the problem of inducing effort exertion: in the case in which there is no need to induce effort, the principal simply monitors in the nonbureaucratic way.

B. *Giving Up on Incentives*

So far I have characterized equilibria that 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 said nothing about how the principal chooses whether to induce effort exertion. This trade-off offers little that is 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 that is surprising here: as a result of lower returns to effort or greater costs of distorted monitoring and decision making, the principal is less likely to induce effort exertion. Therefore, 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 in which the costs of monitoring are quadratic, $\kappa(\rho) = \rho^2/2$; 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 since (14) is always violated, and so monitoring probabilities must be distorted if agents exert effort. First consider the equilibrium propensity to monitor after a complaint. The ex post optimal probability of monitoring is one ($= \kappa'^{-1}(1)$), so any deviation from always investigating after a complaint illustrates the oversight distortions described above. Figure 1 plots the equilibrium propensity to monitor after a complaint as a function of the quality of the bureaucrat's information.[10]

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 one. The reason is that the principal does not induce effort exertion and so investigates efficiently from proposition 3. 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 percent, but this converges back toward the efficient level as the agent becomes more certain. Figure

[10] In this case, the surplus from inducing no effort is 11/16, and the surplus from inducing effort, at the cost of distorted monitoring, is given by

$$q(1) + \frac{1 - q(1)}{4[1 - q(1) + q(1)^2]}.$$

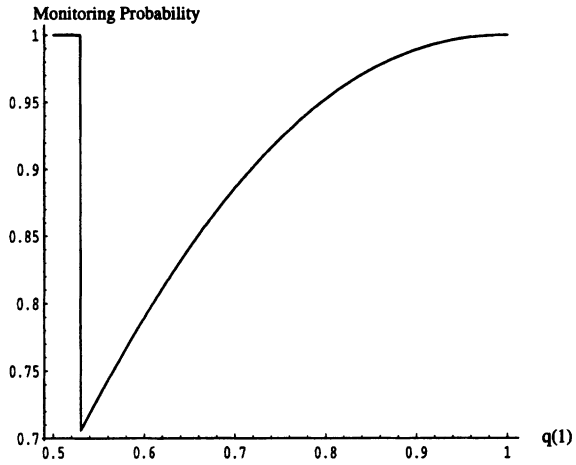

FIG. 1.—Monitoring after a complaint

2 provides analogous information on the propensity to monitor when the consumer is allocated the asset, and so there is no complaint.

Here the ex post efficient monitoring probability is $1 - q(1)$, the probability of an error, and is given by the dashed 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 $1 - q(0) = \frac{1}{2}$ (since 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 monitoring converges to its efficient level as the bureaucrat becomes more certain.

IV. Extensions

A. *The Frequency and Efficiency of Bureaucratic Decision Making*

The results above imply that oversight in bureaucracies is not targeted ex post efficiently. A natural question is how this pattern of ill-focused monitoring in bureaucracies affects the average likelihood of intervention and the number of mistakes that are corrected. In order to do these comparative statics, I consider the quadratic cost case introduced above. Here I show that a characteristic of bureaucracies is that, on average, (i) they monitor more, but (ii) they correct fewer mistakes, than with nonbureaucratic oversight.

PROPOSITION 4. Assume that $\kappa(\rho) = \rho^2/2$, $V(1, 1) = V(1, 0) = v$, and $V(0, 0) = V(0, 1) = 0$. Then bureaucratic oversight $\hat{\rho}$ results in (i)

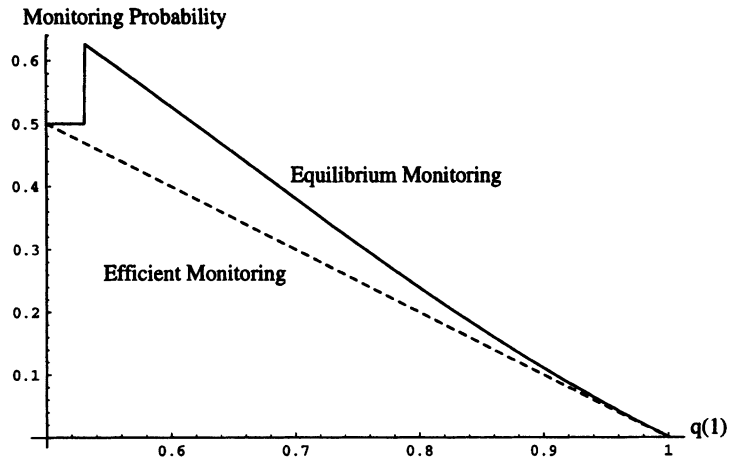

FIG. 2.—Monitoring when the consumer is given the benefit

higher average probability of monitoring but (ii) lower average probability that a mistake will be corrected than nonbureaucratic oversight ρ^{**}.

Proof. See the Appendix.

B. *Consumer-Biased Oversight*

Another 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. But each of these harms the consumer relative to the second-best oversight levels. The ex ante probability that the consumer receives the benefit 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 that monitor in bureaucratic ways deny the asset to the customer more than when monitoring is second-best optimal. Thus bureaucratic oversight is, on average, biased against consumers in that they are more likely to be denied the asset than in the second-best allocation.

C. *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. 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 less likely to be found wrong if she denies the consumer. This results in more delayed decision making in bureaucracies.

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

Define the technologically efficient delay in decision making as $t^*(e)$. This is the delay that would arise if there were no truth-telling concerns and trades off the improved quality of the information against the cost of delay in coming to a decision. The truth-telling condition (relevant only if $e = 1$) analogous to (14), where now effort, oversight, and delay are set at their optimal levels, is now given by

$$[1 - q(1, t^*(1))]\kappa'^{-1}(1) \leq q(1, t^*(1))\kappa'^{-1}(1 - q(1, t^*(1))). \qquad (15)$$

When (15) 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 5. Assume that $V(1, 1) = V(1, 0) = v$ and $V(0, 0) = V(0, 1) = 0$ and that (15) is violated. The optimal delay $\hat{t}(e)$ has the following features: (1) If the agent exerts effort, $\hat{t}(1) > t^*(1)$. (2) If the agent exerts no effort, $\hat{t}(0) = t^*(0)$.

Proof. See the Appendix.

Thus, when the principal induces effort exertion from the bureaucrat, there are what appear to be excessive delays when truth-telling is a binding issue. This arises because the principal is faced with a number of costly instruments (delay and monitoring) that can be used 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.

V. Consumer Choice: Why Are Bureaucracies Always Inefficient?

Thus far, I have addressed problems that arise when bureaucrats are given the power to allocate goods to consumers. But there is an alternative: to allow consumers to play a greater role in allocating goods to themselves, which I address in this section. To allow consumers a greater role in allocating goods, I change the game from that outlined in Section II. One possible way of doing this would simply be to allow consumer choice to be sovereign: their decisions are never overturned. Rather than do this, I reverse the roles of the bureaucrat and consumer in this section. In the model above, the bureaucrat proposes an allocation, the consumer sends a message, and the principal decides an intervention probability. The default if there is no investigation is the allocation proposed by the bureaucrat. (This is what is meant by the allocation of the good by the bureaucrat.) Here I simply reverse the roles of the consumer and bureaucrat: the consumer proposes an allocation, the bureaucrat sends a message, and the principal decides whether to investigate. Note that this encompasses the model of consumer choice, where the consumer is sovereign, since the principal could choose to ignore all messages by the bureaucrat and never intervene to change his decision. More formally, however, all that has changed in the modeling is that the default is now the choice of the consumer. (This is what is meant by allocation of the good by the consumer.) Proposition 6 identifies a limit on the use of bureaucracies.

PROPOSITION 6. If $V(i, i) \geq V(i, j)$, the good is allocated by consumer choice.

The proof of this proposition is obvious and is omitted. It arises because consumers know their own preferences better than bureaucrats do, and so the first-best arises by giving them the discretion to allocate the good. In this case, the sovereign consumer is the optimal allocation. But remember that the condition that generates the optimality of consumer choice ($V(i, i) \geq V(i, j)$) is the condition that generates the absence of distortions in proposition 1! As a result, the only cases in which bureaucracies may be seen are those in which they exhibit the constraints of Section III. To see that bureaucracies are used in these instances, see proposition 7.

PROPOSITION 7. Assume that $V(1, 1) = V(1, 0) = v$ and $V(0, 0) = V(0, 1) = 0$. Then the good is allocated by the bureaucrat.

Proof. See the Appendix.

There are two reasons for this result. First, the consumer always allocates himself the benefit ($a = 1$) so that the principal cannot use his information to focus oversight. Bureaucratic allocation uses such information and so dominates allowing the consumer to choose. Second, consider the outcome when no investigation occurs. If the consumer

allocates the good, he always consumes the benefit. On the other hand, if the bureaucrat allocates the benefit, the default allocation is her belief about what should have been done (α_a). Although this is not perfect, it is at least correlated with social surplus (as $q \geq \frac{1}{2}$) and so yields higher returns to the principal.[11]

Given the discussion above, propositions 6 and 7 are not surprising. Yet they hold the essence of the key insight of the paper, namely, that bureaucracies are used only in instances in which agency problems are hard to solve. Furthermore, the features that lead to the use of bureaucracies also cause their tendency to ignore consumer complaints, overmonitor when consumers are given benefits, delay decision making, and bias oversight against consumers.

VI. Where Should Bureaucratic Inefficiencies Be Seen?

The purpose of this work is not to justify all bureaucratic inefficiency. Instead, my argument is that these inefficiencies should arise only when consumer preferences cannot be aligned with those of the principal. In this section, I consider instances in which this is likely to be so. Up to this point, I have treated the preferences of consumers as exogenous and simply posited that inefficient-looking bureaucracies arise when the preferences of consumers cannot be trusted. But the returns that consumers receive are endogenous in many instances, so that the use of bureaucracies ultimately depends on the cost of aligning the incentives of consumers with those of the principal. I consider a series of reasons why these costs may be prohibitively high.

A. *Pricing*

One way of aligning the incentives of consumers is to charge them for allocations that they like: in this way, beneficial allocations are charged more, and so consumers have less incentive to allocate them inefficiently. For example, in a health care setting, the desire of patients to have tests carried out on them could potentially be reduced by requiring them to contribute to their health care costs. From the perspective of this paper, bureaucracies should then be limited to those cases in which the use

[11] There is one important assumption that underlies this result on the optimality of bureaucrats: the costs of type 1 and type 2 errors are similar and each state is equally likely. But in instances in which there are large costs associated with denying the consumer incorrectly, it may be optimal to allow the consumer to choose the allocation, even if he cannot be trusted. For instance, citizens have the right to expect a police officer or firefighter to arrive if they call with a 911 alarm. Even though their interests may not be in line with those of society, the cost of failing to satisfy a legitimate call makes the optimal default be that the consumer can decide.

of such pricing is inefficient. I consider two such constraints on pricing: (i) incomplete insurance and (ii) harmed incentives.

Insurance.—One instance in which it may not be efficient to charge consumers for goods is one in which there are insurance considerations. For example, consider a case in which a consumer seeks insurance over his health status and could be provided with some costly tests. Suppose that these tests should be carried out only if the patient's condition is serious, on which the consumer has relevant information. In a previous draft of this paper (available on request), I provided a model of this in which I showed that there are two possibly efficient ways of allocating health care in this setting: (i) let the consumer decide what tests he should have or (ii) allow a bureaucrat to make this decision. Assume that the patient is given the right to choose. If the tests are costless to the consumer, he will always have them done, even in instances in which they are not warranted. I show that if the consumer is given this choice, he should be charged a co-payment that is large enough to induce him to select the treatment only if his condition is serious.

While this co-payment allocates treatment efficiently, it reduces risk sharing. Consequently, the drawback of consumer choice is the cost of appropriately aligning incentives, namely the ex ante cost of incomplete insurance. The alternative is not to charge for health care, but to use a bureaucrat to allocate it. This has the inefficiencies described in Section III but improves risk sharing. 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. By contrast, if the ex ante cost of the co-payment is sufficiently large, the bureaucrat should allocate the good. Hence, the demand for insurance generates the way in which the good is allocated.

Incentives.—What matters for the existence of bureaucratic responses is that consumers make rents from the allocation. But moral hazard agency problems for consumers must have this feature to induce efficient actions by consumers; in the absence of ex post rents, there is no incentive to exert effort. To take a trivial example, could professors price grades to students in such a way that students efficiently choose their own grade? In this world, a student would pay a higher price for an A than for a B. Or a suspect would pay a fee not to be arrested. But there is a serious problem with allowing such pricing in that charging people for higher grades reduces incentives, since the value of a better grade now falls. Equally, making suspects closer to indifferent about whether they are arrested harms incentives to deter crime. As a result, bureaucrats must be used to allocate assets in most incentive situations, resulting in the problems described above.

B. Unobserved Allocations

Most of the examples above concern the allocation of public-sector goods, such as arrests, entry of immigrants, and so on. I have emphasized these goods because even when allocations are easily observed, it is very difficult to observe surplus without investigation. These issues are especially stark in instances in which consumers are allocated goods at no cost. However, there are also market relationships in which these issues become relevant, namely, those in which it is difficult to observe the precise nature of the transfer between a consumer and an employee. To phrase this another way, the message of the last subsection is that sometimes prices can be used to induce consumers to reveal information honestly. (For instance, charging consumers for health care may make their information more credible.) But this relies on the assumption that it is relatively easy to observe the goods that are being given to the consumer, which can then be compared to the price they paid for them. But this is not always so, even in the private sector.

To see this, remember the truth-telling constraint above in (11): the incentive to capitulate to consumers becomes tempting when the cost of observing surplus in that state ($\kappa(1, n)$) is sufficiently high. In other words, when it is simple to identify the nature of the transfer between agent and consumer, and this is all that is necessary to know surplus, these problems are mitigated. For example, suppose that I purchase a Toyota but the dealer gives me a Mercedes to keep me happy. Whether this results in the kinds of bureaucratic problems above depends on the observability of the price-quantity transfer to the consumer. In the example above, inventory can easily be checked against the price paid by the consumer; in other words, $\kappa(1, n)$ is low. Consequently, there is little need to distort oversight.

But there are other market relationships in which agents can allocate benefits to consumers with little likelihood that this is seen. (What matters here, of course, is not that consumers get unobserved benefits, but that the principal would prefer that these benefits not be given.) One such private-sector example is the behavior of employees in cinemas, who are supposed to enforce rules that underage patrons cannot see R- or NC-17-rated movies. Although it is easy to observe that the customer has paid for a ticket, it is difficult for a principal to easily observe the customer's age. This example easily satisfies the criteria for the model to be relevant, and indeed, underage customers are often admitted to movies that they should not attend. As surplus is hard to see in this instance, the incentive to simply give the consumer what he wants becomes relevant even in this market environment.

VII. Conclusion

Bureaucrats certainly get bad press. Disparaging the contributions of pen pushers and bean counters has become part of private discourse, media reports, and political campaigns. There is little doubt that bureaucracies have problems for a variety of reasons. For example, in many public-sector settings, wages cannot be changed by superiors, employees are promoted by seniority, firing is next to impossible, and objectives for superiors are impossibly vague. This paper offers a more benign view of the supposed inefficiencies of bureaucrats, with limitations on their ability to allocate resources 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. In this sense, the ideas of this paper harken back to Coase (1937), who argues that the allocation of activities into markets and firms is not random, but rather is part of a process designed to minimize transaction costs. Much of the literature on the theory of the firm has used this insight to show how the nature of activities within firms often looks very different from those done through markets. Similarly, this paper proposes that activities are not randomly distributed between consumer choice, a close cousin of markets, and bureaucracies and that the features that allocate activities to bureaucracies ultimately limit their efficiency.

The central difficulty with bureaucrats is that they work in settings in which there are no reliable output measures on which to base pay. Given this problem, superiors rely on customer complaints to focus their attention since customers are often informed of mistakes, which can then be corrected. The key problem with bureaucracies emphasized here is that bureaucrats tend to be used only in those situations in which consumers cannot be fully trusted to reveal errors by bureaucrats, which limits their efficiency. 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. In other work in this area, I have collected evidence on the behavior of police officers to suggest that these incentives are indeed empirically important. Specifically, the Los Angeles Police Department changed its oversight of officers in 1998; an administration that was historically reluctant to investigate consumer complaints and penalize officers for infractions radically changed course. Since 1998, the policy of the department has been both to more seriously investigate all consumer complaints against officers and to fire or suspend officers who have been found guilty of infractions. This has resulted in a huge increase in consumer complaints, and over 10 percent of all officers have been suspended or fired since January 1998 (compared to an annual rate of less than 0.05 percent before then). In

Prendergast (2002*b*), I show that this has coincided with a large reduction in the arrest rates of police officers (e.g., narcotics arrests were down roughly 50 percent by 2001) and a noticeable reduction in the willingness of officers to confront suspects (e.g., by using force). As a result, this incentive to capitulate to consumers (in this case suspects) to avoid oversight does appear to be empirically relevant.

The main results of the paper are concerned with overcoming this incentive. 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 little scrutiny of particularly incompetent bureaucrats after a complaint. The implies a bureaucracy that 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 Dewatripont, Jewett, and Tirole [1997] and Dixit [2002]), it does at least paint a picture of bureaucratic life that seems more realistic than that offered by the standard agency model. Closest in spirit to this approach is the paper by Banerjee (1997), who considers the use of incentive contracting to constrain bureaucratic corruption. As in my model, penalizing bureaucrats for incorrect decisions plays a role in inducing bureaucrats to make better decisions, but at the cost of distortionary activities. In Banerjee's model, these costs take the form of excessive red tape, whereas here the harmful activity is the temptation to capitulate to consumers.

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 allow only one party, the consumer, to complain. In some situations, this is unrealistic because other aggrieved parties could also complain about a mistake in the consumer's favor. For example, the victim of a crime could complain about the failure of an officer to arrest a suspect. As such, I think that this paper is most relevant to those cases in which 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, 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 in which there are clear and well-informed parties on each side of the transaction, where a mistake in either direction will result in a complaint.

Note also the *assumption* that bureaucrats wish to avoid investigation. An alternative assumption would allow wages to depend more generally

on perceptions of ability, such as in a career concerns models. In this model, ability would reflect the capacity to correctly identify the right treatment. Consider a scenario in which agents are rewarded in a linear fashion on the probability that they identify the treatment correctly. Then wages rise if the agent is proved to be correct and fall if incorrect in a symmetric fashion. This is the environment identified in Prendergast (2002*a*). I show that when there is limited observability of cases in which agents are not investigated, it remains the case that there is an incentive to avoid investigation despite the inherent symmetry of rewards. The reason for wishing to avoid investigation is that because consumers point out errors, *on average,* an investigation reveals harmful information on the agent. As a result, they wish to avoid investigation, all else equal, and the qualitative results of the paper continue to hold.

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 III. Thus observed bureaucracies must be inefficient. In other words, bureaucracies work better when consumers complain efficiently (as in Sec. II), 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 in which $e = 1$. On the condition that truth-telling is satisfied, 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 (11) holds. The objective of the principal is then to

$$\max_{\rho(a,m)} \frac{1-q(1)}{2}[\rho(0,\ c) - \kappa(\rho(0,\ c))] - \kappa(\rho(0,\ n)) + \tfrac{1}{2}\{[1-q(1)]\rho(1,\ n) - \kappa(\rho(1,\ n))\} \tag{A1}$$

subject to (11). When (11) is violated at the optimal monitoring propensities, the principal chooses monitoring propensities such that (11) binds. By substi-

tution, the objective function is

$$\max_{\rho(a,m)} \frac{1-q(1)}{2}[\rho(0, c) - \kappa(\rho(0, c))] - \kappa(\rho(0, n)) + \frac{1}{2}\left\{\frac{[1-q(1)]^2}{q}\rho(0, c) - \kappa\left(\frac{1-q(1)}{q}\rho(0, c)\right)\right\}. \quad \text{(A2)}$$

Straightforward differentiation yields

$$\hat{\rho}(0, n) = 0 \quad \text{(A3)}$$

and

$$q(1)\left\{[1 - \kappa'(\rho(0, c))] + [1 - q(1)] - \kappa'\left(\frac{1-q(1)}{q}\rho(0, c)\right)\right\} = 0. \quad \text{(A4)}$$

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

$$q(1)\left\{\kappa'(\hat{\rho}(0, c)) + [1 - q(1)]\kappa'\left(\frac{1-q(1)}{q}\hat{\rho}(0, c)\right)\right\} = q(1) + [1 - q(1)]^2. \quad \text{(A5)}$$

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

$$\kappa'\left(\frac{1-q(1)}{q}\hat{\rho}(0, c)\right) > 1 - q(1).$$

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 monotonically increasing in monitoring 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 (14) is violated. This implies that $q(1)/[1 - q(1)] = \hat{\rho}(0, c)/\hat{\rho}(1, n)$. 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 in which $e = 0$. 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

$$\max_{\rho(a,m)} \frac{1-q(0)}{2}[\rho(0, c) - \kappa(\rho(0, c))] - \kappa(\rho(0, n)) + \tfrac{1}{2}\{[1 - q(0)]\rho(1, n) - \kappa(\rho(1, n))\} \quad \text{(A6)}$$

subject to the salary w_0 meeting the agent's outside opportunities. This implies that monitoring propensities are given by $\kappa'(\hat{\rho}(0, c)) = 1$, $\hat{\rho}(0, n) = 0$, and $\kappa'(\hat{\rho}(1, n)) = 1 - q(0)$, 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

In the quadratic cost case, it is simple to show that if the agent is induced to exert effort, the allocations are given by

$$\hat{\rho}(1, n) = \frac{1 - q(1)}{1 - q(1) + q(1)^2} \geq 1 - q(1) = \rho^{**}(1, n) \tag{A7}$$

and

$$\hat{\rho}(0, c) = \frac{q(1)}{1 - q(1) + q(1)^2} \leq 1 = \rho^{**}(0, c). \tag{A8}$$

Thus oversight is "too low" after a complaint and "too high" without a complaint. The average probability of monitoring is then given by

$$\tfrac{1}{2}\hat{\rho}(1, n) + \frac{1 - q(1)}{2}\hat{\rho}(0, c) = \frac{1 - q(1)^2}{2[1 - q(1) + q(1)^2]}.$$

But if consumers point out all errors, as in Section II, the probability of an investigation is given by

$$1 - q(1) < \frac{1 - q(1)^2}{2[1 - q(1) + q(1)^2]}.$$

Thus bureaucracies monitor more on average.

Mistakes occur with probability $1 - q(1)$, so the fraction of mistakes that would be corrected in the absence of distorted monitoring is

$$\frac{1 - q(1)}{2}\rho^{**}(0, c) + \frac{1 - q(1)}{2}\rho^{**}(1, n) = \frac{[1 - q(1)][2 - q(1)]}{2}.$$

When truth-telling problems imply that monitoring is given in (A7) and (A8), the analogous correction rate is

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

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

Proof of Proposition 5

First consider the case in which $e = 1$. The objective of the principal is to

$$\max_{\rho(a,m),\tau} e^{-rt}\Big\{q(1, t) + \frac{1 - q(1, t)}{2}[\rho(0, c) - \kappa(\rho(0, c))] - \kappa(\rho(0, n)) + \tfrac{1}{2}\{[1 - q(1, t)]\rho(1, n) - \kappa(\rho(1, n))\}\Big\} \tag{A9}$$

subject to (15). When (15) 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 (15) be given by λ, and let

$$\Gamma = \frac{1 - q(1, t)}{2}[\rho(0, c) - \kappa(\rho(0, c))] - \kappa(\rho(0, n)) + \tfrac{1}{2}\{[1 - q(1, t)]\rho(1, n) - \kappa(\rho(1, n))\}$$

be the social surplus from the allocation at the time of delivery. Then the optimal choice of timing to make a decision is given by

$$r\Gamma = \frac{dq(1, t)}{dt}\left\{1 - \frac{\rho(0, c) - \kappa(\rho(0, c))}{2} - \rho(1, n) + \lambda[\rho(0, c) + \rho(1, n)]\right\}. \quad \text{(A10)}$$

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

$$r\Gamma = \frac{dq(1, t^*)}{dt}\left[1 - \frac{\hat{\rho}(0, c) - \kappa(\hat{\rho}(0, c))}{2} - \hat{\rho}(1, n)\right], \quad \text{(A11)}$$

where $\kappa'(\hat{\rho}(0, c)) = 1$, $\hat{\rho}(0, n) = 0$, and $\kappa'(\hat{\rho}(1, n)) = 1 - q(1, t^*)$. Diminishing returns in $\Omega(q)$ and $q_2(e, t^*(e))$ guarantee uniqueness. But when the truth-telling constraint binds, the shadow price $\lambda > 0$, and hence $\hat{t} > t^*$ from (A10).

Next consider the case in which $e = 0$. 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 (A11), except $q(0, t)$ substitutes for $q(1, t)$.

Proof of Proposition 7

When the consumer allocates the good, he always chooses $a = 1$ (since the principal does not always reverse inefficient decisions). As a result, the consumer's information is not relevant for setting oversight levels, and so oversight is based solely on the agent's information. If the agent agrees with the consumer's allocation and the truth-telling condition does not bind, the principal investigates with probability

$$\rho_1(1, n) = \kappa'^{-1}(1 - q(e)); \quad \text{(A12)}$$

if he disagrees, the principal investigates with probability

$$\rho_1(1, c) = \kappa'^{-1}(q(e)). \quad \text{(A13)}$$

The first argument is the allocation made by the consumer (always one), and the second is based on whether the agent disagrees (c) or not (n). In the absence of an investigation, the allocation is one.

Note that with these oversight levels, the truth-telling condition is not violated. The relevant truth-telling condition with high effort is now given by

$$[1 - q(1)]\kappa'^{-1}(q) \leq q(1)\kappa'^{-1}(1 - q(1)), \quad \text{(A14)}$$

which is violated only if

$$\frac{1-q(1)}{q(1)} > \frac{\kappa'^{-1}(q(1))}{\kappa'^{-1}(1-q(1))}. \tag{A15}$$

But this cannot be true by the convexity of marginal cost function $\kappa''' \geq 0$. This generates the optimal contract offered with consumer oversight in this case, where $e = 1$ arises by appropriate choice of Δ, since the truth-telling condition does not bind.

Now compare this to bureaucratic allocation, where the principal chooses to ignore all messages by the consumer and investigates with probability $\rho_1(1, n)$ if allocation $a = 1$ is proposed by the bureaucrat and $\rho_1(1, c)$ if allocation $a = 0$ is proposed. This replicates the oversight under the consumer choice outcome. As a result, if there is an investigation, it carries the same returns to the principal as with consumer choice and similarly satisfies truth-telling. But there are two reasons why bureaucratic oversight dominates this optimal allocation with consumer choice.

First, the principal strictly prefers to offer oversight different from $\rho_1(1, n)$ and $\rho_1(1, c)$: instead, he chooses oversight using the consumer's information, as given by ρ^{**} in Section III. This is true for any $q \geq \frac{1}{2}$. Second, when oversight is held fixed, there is a return to allowing the bureaucrat to choose the default. With consumer choice, the default is that $A = 1$ if the principal does not investigate. Thus the return to the principal in states in which he did not investigate is

$$\frac{q[1-\rho_1(1, n)] + (1-q)[1-\rho_1(1, c)]}{2}.$$

By contrast, with these oversight propensities, if the bureaucrat allocates the good, the return to the principal in states in which he did not investigate is

$$\frac{q[1-\rho_1(1, n)] + q[1-\rho_1(1, c)]}{2} \geq \frac{q[1-\rho_1(1, n)] + (1-q)[1-\rho_1(1, c)]}{2}.$$

Therefore, the bureaucratic allocation is preferred because $q(1) > \frac{1}{2}$.

References

Banerjee, Abhijit V. "A Theory of Misgovernance." *Q.J.E.* 112 (November 1997): 1289–1332.

"Besmirched Deportland Wrestles with the INS." *New York Times* (August 31, 2000), p. A19.

Cannon, Lou. "One Bad Cop." *New York Times Mag.* (October 2, 2000).

Coase, Ronald H. "The Nature of the Firm." *Economica* 4 (November 1937): 386–405.

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

de Figueiredo, Rui J. P., Jr.; Spiller, Pablo T.; and Urbiztondo, Santiago. "An Informational Perspective on Administrative Procedures." *J. Law, Econ., and Organization* 15 (April 1999): 283–305.

Dewatripont, Mathias; Jewitt, Ian; and Tirole, Jean. "Missions and the Accountability of Government Agencies." Manuscript. Toulouse: Univ. Toulouse, 1997.

Dixit, Avinash. "Incentives and Organizations in the Public Sector: An Interpretative Review." *J. Human Resources* 37 (Fall 2002): 696–727.

Friebel, Guido, and Raith, Michael. "Abuse of Authority and Hierarchical Communication." Manuscript. Chicago: Univ. Chicago, Grad. School Bus., 2001.

Leaver, Claire. "Squawk and Capture: Influencing Appointed Regulators." Manuscript. Bristol: Univ. Bristol, 2001.

McCubbins, Matthew D., and Schwartz, Thomas. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American J. Polit. Sci.* 28 (February 1984): 165–79.

Prendergast, Canice. "Consumers and Agency Problems." *Econ. J.* 112 (March 2002): C34–C51. (*a*)

———. "Selection and Oversight in the Public Sector, with the Los Angeles Police Department as an Example." Manuscript. Chicago: Univ. Chicago, 2002. (*b*)

Wilson, James Q. *Bureaucracy: What Government Agencies Do and Why They Do It.* New York: Basic Books, 1989.