

A THEORY OF MISGOVERNANCE*

ABHIJIT V. BANERJEE

This paper tries to explain why government bureaucracies are often associated with red tape, corruption, and lack of incentives. The paper identifies two specific ingredients that together can provide an explanation: the fact that governments often act precisely in situations where markets fail and the presence of agency problems within the government. We show that these problems are exacerbated at low levels of development and in bureaucracies dealing with poor people. We also argue that we need to posit the existence of a welfare-oriented constituency within the government in order to explain red tape and corruption.

I. INTRODUCTION

I.A. Goals of a Theory of Misgovernance

The stereotypical view of government bureaucrats, as articulated in the press for example, is that they are lacking in incentives, obsessed with red tape, and probably corrupt. The point of departure of this paper is that while such views may well be correct, it is worth understanding to what extent these phenomena can be explained without departing from the standard paradigm where the government is a benevolent social planner. In other words, we are looking for an explanation of government failures that makes no reference to the rapacity of governments, their monopoly of state power, or the unique sociological status of governments.¹

To pose the problem in this way is not to deny that some

*This paper was inspired by many conversations with Andrei Shleifer. Two anonymous referees made extremely helpful comments. I have also profited from comments by Daron Acemoglu, Andres Almazan, Dipak Banerjee, Tuli Banerjee, Gary Becker, Douglas Bernheim, Christopher Clague, Peter Diamond, Avinash Dixit, Drew Fudenberg, Oliver Hart, Patrick Legros, Eric Maskin, Andrew Newman, Thomas Piketty, Rohini Somanathan, Jean Tirole, Jorgen Weibull, and seminar participants at Princeton University, the University of British Columbia, the Kennedy School of Government at Harvard University, the University of Chicago, and Harvard University. Some of these arguments were in the notes that were circulated some time ago as "The Costs and Benefits of Corruption." This work was carried out when the author was an IPR junior fellow. It was supported by financial assistance from IRIS and the National Science Foundation. However, the views expressed here are strictly the author's own. The author also acknowledges the hospitality of DELTA in Paris where this work was started.

1. Theories of the government failures based on the government's rapacity and its monopoly of state power abound. Perhaps the most articulate statement is to be found in the works of Mancur Olson and his followers (see, for example, Olson [1993]). A very different theory of government failures which emphasizes the unique sociological status of modern governments and the consequent limits on what the government can and cannot do, is in Wilson [1989]. See also the formalization of the Wilson's ideas in Dixit [1996].

governments are extremely rapacious. Nor is it to deny that the sociological status of governments is both important and interesting. But it is to emphasize that a significant part of what we see as government failures may exist even when a government has the best of intentions and is subject to no special sociological constraints.

To overlook this simple point runs the danger, in our view, of limiting our understanding of where and under what circumstances governments perform relatively well and therefore biasing our policy stances. To take a simple instance, if we observe a high degree of corruption in a particular government bureaucracy and assume that all other bureaucracies in the same government will be equally corrupt, we may recommend against specific forms of government activism that may in fact work well.

The basic claim of this paper is that it is possible to develop a theory of misgovernance by a benevolent government based on two eminently reasonable premises: one, that a substantial part of what governments do is to respond to market failures; and two, like all other organizations, the government has agents who are more interested in their own welfare than in any collective goals. And, perhaps more importantly, the theory set up for the sake of this explanation has sensible and useful implications about the performance of different government bureaucracies under different circumstances.

The model we set up is extremely simple. There are three types of agents: the government, bureaucrats, and the people outside. The government in our model has a set of publicly provided private goods that the people want. It is interested in allocating them in a way that maximizes social welfare. These goods may be educational opportunities; beds in hospitals; licenses to produce, import, or pollute; or even irrigation water.² To avoid being unnecessarily specific, we will just call them slots.

These slots are scarce in the sense that the number of people who want them exceeds the number of goods. Not all the people who want these slots value them equally; we assume that there are two types of which one has a higher willingness to pay for the slots. Clearly, in an efficient allocation people of this type should get the slots ahead of the others, and we would typically expect the market to deliver this outcome. However, here this is not nec-

2. Wade [1982] provides a fascinating description of the process of allocation of irrigation water (in Southern India) by a public bureaucracy.

essarily the case because we assume that at least for some people, the willingness to pay is higher than the ability to pay. The obvious reason for such a discrepancy would be a credit market imperfection, but it could also arise out of a labor market imperfection that limits the number of hours someone can work (most jobs actually do this to a greater or a lesser extent).

This assumption of a capital market imperfection is relatively uncontroversial in the context of education or health. It is less obvious that those who are bidding for trade or production licenses are generally credit-constrained, but it may not be unreasonable to assume that this constraint binds for at least some of them. Certainly in the early years of development planning, limited and unequal access to credit was often the stated justification for the licensing of industrial production, imports, exports, and access to foreign exchange.³

The fact that the market may fail to allocate the slots efficiently is going to be key to our model. It explains both why the government is involved in the allocation of these goods as well as why imitating the market will not be the best way to allocate them.

In our model the actual allocation of the slots is the responsibility of a bureaucrat. We assume that the bureaucrat cannot observe the value put on a slot by each person who demands it. We also assume that the bureaucrat cares only about his own welfare and that government cannot perfectly monitor the mechanism used by the bureaucrat to allocate the slots. Therefore, there are really two potential incentive problems: the applicants for the slots may lie to the bureaucrat about their willingness to pay, and the bureaucrat may lie to the government about the mechanism he is using.⁴

As we will show, the combination of these quite elementary assumptions yields a model that has a rich set of predictions:

3. While this form of government intervention eventually proved to be a constraint on development and was probably based on an excessive mistrust of the price system, there is little reason to believe that the arguments in their favor were disingenuous. In other words, the eventual abandonment of these systems does not imply that the initial decision to adopt them was not *ex ante* social welfare maximizing, given the information and the understanding that the government then had.

4. The mechanism design problem that the bureaucrat solves is of some independent interest. There is now a growing literature on general mechanism design problems with credit-constrained agents. See, for example, Aghion and Burgess [1993], Bolton and Roland [1992], Che and Gale [1994], and Lewis and Sappington [1996]. Our paper departs from these in emphasizing the role of red tape in designing such mechanisms.

first, it can explain why bureaucrats will want to use red tape, interpreted as completely pointless bureaucratic procedures that one has to endure in dealing with bureaucracies. Second, the model can explain corruption. Here it is worth emphasizing that in order to explain corruption one needs to explain more than moneymaking by government bureaucrats: one needs to explain *illegal* moneymaking. And to do so, one needs to explain why the government makes it illegal to make money.⁵ Third, the model explains why, under certain circumstances, the government will give bureaucrats very low powered incentives or no incentives at all.

At a very different level the model also allows us to ask what would change if the government were interested in making money rather than in social welfare. It turns out that in this case there would be no red tape at all, unless there were unobservable differences in the ability to pay and even when there are such unobservable differences, there will be *less* red tape in this case than in the case where the government is welfare-minded. The same is true of corruption: there would be no corruption in the world of this model if the government did not care about social welfare. In other words, the assumption that the government is rapacious makes it harder to explain red tape and corruption. This is less paradoxical than it appears: as will be explained in the following pages, both corruption and less obviously, red tape, arise out of the government's efforts to control the bureaucrat in the social interest. If the government did not have society's interest at heart, there would be no need to have such controls.

It is also worth asking whether the assumption of agency problems within the government is necessary for our specific results. To check this, we also consider the case where both the government and the bureaucrat are welfare-minded. We show that in this case there will be no red tape and (obviously) no corruption. In other words, a conflict of interest within the government is key to our story.

Finally, the model gives a number of predictions about the determinants of red tape and corruption. In particular, we show that, on the whole, red tape and corruption are more likely to arise where ability to pay is low relative to the willingness to pay,

5. This insistence on explaining why the government makes corruptible rules is what distinguishes our framework from much of the existing literature on corruption (see, for example, Shleifer and Vishny [1993]).

where the goods being allocated are particularly scarce and where there is inequality in the ability to pay. We also find that it is precisely in these environments that bureaucrats may face weak incentives. We interpret these as saying that government failures are most likely in bureaucracies dealing with poorer section of society and in poor countries.⁶

We postpone providing intuition for these results until we have presented the key ingredients of the model. This is the subject of the next subsection. Once the model is presented, we will present some relatively loose analysis that will explain the basic properties of the model and provide intuition for the results claimed above. More formal analysis is provided in the later sections of the paper.

I.B. The Model

We assume that the set of slots being allocated is of Lebesgue measure 1 and the population of applicants to be of Lebesgue measure $N > 1$. The applicants can be of two types, L and H , or alternatively low and high. The low type generates a return L if awarded the slot, while the high type generates a return of H . We assume that these are both the social and private returns and that $L < H$. We assume that the fraction of type H applicants is $N_H < 1$ and that of type L is N_L . Finally, we assume that the applicants are risk-neutral and have quasi-linear preference over slots and money; i.e., if an applicant gets a slot worth H with probability π and pays an amount p_H for it his net utility will be $\pi H - p_H$.

The applicants for the slots are cash-constrained in the sense that their valuation of these slots may exceed their ability to pay for them. We model the cash constraint as an upper bound, y , on each applicant's ability to pay. We do not allow the government to relax this constraint by giving people money on the grounds that if the government started giving away money a lot of people may claim that they want a slot in order to get the money. In this

6. This is consistent with the evidence presented in Mauro [1995] about the correlation between government failures and level of development. We are aware, of course, that there are other reasons why bureaucrats in poorer countries are corrupt. For example, the salaries paid to responsible government servants in many LDCs do not seem to be commensurate with their responsibilities. In other words, it is possible that the bureaucrats in these countries are corrupt because they get paid less than their efficiency wages. However, this begs the question of why the government sets salaries that are so low. Our model has the advantage of giving reasons for why the government may choose to let the bureaucrat make money.

section and the next two we will assume that y is the same for all applicants. This assumption will be relaxed in Section IV.

The slots belong to the government, but the actual allocation of the slots is the responsibility of a bureaucrat. This distinction between the government and the bureaucrat is central to the argument we make here: in our model the bureaucrat chooses the mechanism that is used for allocating the slots, while the government is responsible for rewarding and punishing the bureaucrat.⁷

Regarding the preferences of our two main actors—the bureaucrat and the government—for most of the paper we make the assumption that the bureaucrat cares only about the total amount of money he makes, less the costs of implementing red tape and any other costs, while the government cares only about social welfare.⁸ These preferences make the most sense if we assume that both the government and the bureaucrat are risk-neutral and face no liquidity constraint. In this case the government can always satisfy the bureaucrat's participation constraint by making him a lump sum transfer, and, on the other side, if the government feels that the bureaucrat is making too much money and wants to recoup some of the revenue from the sale of the slots, all it has to do is to set a fixed fee for each slot. We will, however, also consider what happens if both the government and the bureaucrat are only interested in making money, as well as the case where both are welfare-minded.

The mechanism chosen by the bureaucrat for allocating the slots will typically combine prices and what we call red tape. In other words, an applicant who wants a slot will have to pay a certain amount and also go through a certain amount of red tape before he gets the slot. We model red tape as a pure waste of time.⁹ We assume that going through a unit of red tape costs the

7. The distinction we make here between the government and the bureaucrat parallels the distinction made by Laffont and Tirole [1993] between the constitution-maker and the regulatory agency.

8. The two preferences we have specified are clearly both quite extreme. In reality, a welfare-oriented government may also care about revenue because of budgetary concerns. However, allowing the government to put a small weight on revenue does not change our results. Also the way we have modeled the welfare-oriented preferences assumes that even a welfare-oriented government does not care about how the allocation of the slots affects the distribution of wealth. This is deliberate; allowing the government a more complex objective makes it easier to explain why it might generate inefficient outcomes: our present formulation therefore provides the sharpest test of our theory.

9. Nothing essential would change if we assumed, instead, that red tape actually produces information. Also, despite being a waste of time, screening is an important social function, and therefore we do not interpret the use of red tape per se to be a sign of inefficiency. It is rather the red tape that is in excess of the socially necessary amount that we view as a measure of governmental inefficiency.

applicant δ . These costs may be thought of as the losses in productivity from delays, time costs of waiting in lines, or simply the emotional costs of being harassed. We will assume that this is a nonmonetary cost in the sense that having to bear it does not reduce the applicant's ability to pay.¹⁰ We also assume that the cost per unit of time to the bureaucrat of inflicting red tape on an applicant is ν , where ν/δ is small.

To complete the model, we need to specify the ways in which the government can provide incentives for the bureaucrat. For the time being, we will assume that the government does not observe the mechanism used by the bureaucrat to allocate the slots: it observes neither the amount of red tape nor the prices charged by the bureaucrat. This assumption is relaxed in Section III, where we allow the government to punish the bureaucrat for using the wrong mechanism but put a bound on such punishments.

However, we do allow the government the possibility of providing the bureaucrat with some incentives on the basis of how the bureaucrat has allocated the slots that were given to him to allocate. There are several alternative ways of introducing such incentives that give more or less equivalent results. Here we choose a formulation that is analytically convenient at the cost of being somewhat crude. We assume the following.

(i) The government samples a small fraction of those who are given slots by the bureaucrat and determines their types. Because of the assumption that the number of slots forms a continuum, the sample tells the government the exact number of slots that went to type L applicants.¹¹

10. This is more than we really need to assume: our results only require that the wasted time does not reduce the applicant's ability to pay one for one. Interpreted in this way this assumption seems to be quite consistent with our suggested interpretations.

11. This of course requires that the government can tell who are type L applicants. It is legitimate to ask why, if we allow the government access to a technology for determining the type of the applicant, we also do not do so for the bureaucrats. However, the situation we have in mind is one where it is quite costly to directly establish the applicant's type, and therefore a bureaucrat will not want to do so (especially since, as will become evident, there are cheaper ways to screen). On the other hand, we imagine that each bureaucrat allocates many slots, and therefore, if the government can influence the allocation of all these slots by sampling a small fraction of those who get the slots and determining their types, it may very well be worthwhile.

It may also be the case that it is much more difficult to discover the applicant's true type at the time the slots are being allocated than it is in the long run: information has a way of leaking out on its own over time. Since the bureaucrat typically has a long-term relationship with the government, the government may be able to use this information against the bureaucrat much more easily than the bureaucrat can use it against the person who got the slot.

It is also clear that, ideally, all these arguments should be modeled formally, but we do not see any way of doing this without making the paper unreadable.

(ii) The government imposes a fine F on the bureaucrat for each slot in excess of $1 - N_H$ which goes to an L -type applicant, where $1 - N_H$ is both the fraction of slots that would go to type L applicants in the first-best allocation and the minimum fraction of slots that must go to type L applicants in any allocation. In other words, the bureaucrat who gives slots to N'_L type L applicants, pays a total fine of $(N'_L - 1 + N_H)F$.

(iii) We assume that the government gets to choose F , and until Section III we do not impose any bound on how large F can be.

This particular formulation is, admittedly, crude. However, note that while we could allow the government to use more sophisticated incentive schemes, this would not expand the set of implementable outcomes or reduce the cost of implementing them.¹² Intuitively, what matters from the point of view of the bureaucrat's incentives is the marginal cost of giving an additional slot to a type L applicant. In this formulation this marginal cost turns out to be just F , which, by assumption, the government can set at any level it wants.

We also assume that the government can always control the number of slots that the bureaucrat allocates in order to avoid the possibility of an additional monopoly inefficiency that arises because the bureaucrat rations the slots to raise their price. This is an additional complication that is unimportant to our basic line of argument and therefore, we feel, best avoided.

To end the description of the model, the sequencing of the actions is as follows. The government first chooses F . Then, given F , the bureaucrat chooses the mechanism for allocating the slots. The applicants make their choices taking the mechanism as given.

I.C. Some Rudimentary Analysis

In order to understand the logic of our model, we start with a special case where the analysis is extremely straightforward. The bureaucrat in this case is only allowed to charge a price to those who receive the slot. We will call such mechanisms *winner-pay* mechanisms and distinguish them from *all-pay* mechanisms, which are mechanisms where all participants have to pay, irrespective of whether or not they get slots.

12. Strictly, this is only true when all bureaucrats are identical in terms of their preferences, which is true in all sections of the paper except Section III.

Within this special model, first consider a situation where both the bureaucrat and the government are welfare-oriented. In this case, so long as y is not too low, the first-best outcome in which all the high types get a slot and nobody suffers any red tape, can be implemented by using a price mechanism. Essentially all we have to do is offer the low type a sufficient discount on what the high type is paying, and then the low type will be willing to accept the lower probability of getting the good. The only problem arises when y is very low; then it is impossible to give the low type a large enough discount (this is obvious when $y = 0$). We state the precise claim in the following.

CLAIM 1. Under the assumption that the government and the bureaucrat are both social welfare maximizers, the first-best allocation can be achieved if $y \geq L - L(1 - N_H)/N_L$, by using the following allocative mechanism.

If $y > L$, those who declare themselves to be a type H pay a price $p_H = \min(y, H - (H - L)(1 - N_H)/N_L)$ and always get the slot. Those who claim to be type L get the slot with probability $(1 - N_H)/N_L$ and pay a price $p_L = L$ when they get a slot. If $y \leq L$, those who declare themselves to be a type H pay a price $p_H = y$ and always get the slot. Those who claim to be type L get the slot with probability $(1 - N_H)/N_L$ and pay a price $p_L = L - (L - y)N_L/(1 - N_H)$ when they get a slot.

We omit a formal proof of this proposition since it is a simple extension of the verbal argument given in the text.

One practical way to implement the mechanism proposed here is for the government to sell N_H slots at the market price to the type H 's (they are after all paying what would be the market price) and to reserve the rest for allocation to the type L 's at a lower than market price. In fact, the task of allocating slots to the high types may even be given over to the private sector in order to reduce the bureaucratic burden on the government. One observes many examples of such mechanisms in the real world (for example, certain medicines may be sold both on the market and through the public distribution system).

Consider next the other extreme case—where both the government and the bureaucrat are only interested in making money. In this case it is in the government's interest to allow the bureaucrat to freely maximize profits (i.e., to set $F = 0$) and then collect the revenue from the bureaucrat as a lump sum fee (or equivalently, by charging the bureaucrat a high enough price per

slot). Now as long as $y < L$, the maximum profit the bureaucrat can get is y per slot.¹³ This can be achieved by setting a single price equal to y and then offering everybody an equal chance of buying the slot at that price. No red tape will be used. In other words, a purely rapacious government will also avoid red tape (at the cost of generating a poor final allocation).

Finally, let us consider the intermediate case in which there is a conflict of objectives. Given our assumptions, the government can always induce the bureaucrat to give a slot to each high type person—simply by setting F sufficiently high. However, the bureaucrat will not want to use a mechanism of the type described in Claim 1; he makes too little money on the low type. Rather he would want to set the price to both types equal to y (at least as long as $y < L$). However, if both types are paying the same and those who declare themselves to be the high type are getting the slot for sure, everyone will claim to be the high type. To restore incentive compatibility, the bureaucrat will have to threaten anybody who claims to be a high type with enough red tape; i.e., the amount of red tape, T_H will have to satisfy

$$(1) \quad L - y - \delta T_H = (L - y)(1 - N_H)/N_L.$$

This solution will be optimal for the bureaucrat so long as red tape does not cost him too much; i.e., ν is small relative to δ .

This argument assumes that $y < L$. No red tape would arise if $y \geq L$: the bureaucrat could simply charge the type H applicants $p_H > L$ and the type L 's L , and incentive compatibility would be automatic (see Section II for a formal statement of this claim).

Finally, observe that, in the case where $\nu = 0$, for any positive value of F the bureaucrat will use the mechanism described in the previous paragraph and give a slot to every type H while charging both types a price y . Screening will be achieved entirely by the use of red tape. This follows from the fact that by using this mechanism the bureaucrat is getting as much money as he can ever get; every slot is earning the maximum amount y . Therefore, he loses nothing by using red tape to do all the screening (especially since $\nu = 0$, but a similar result holds when ν is close to 0).

13. Since we do not allow him to charge those who do not get the slot.

I.D. What Do These Results Tell Us?

The results in the previous section offer a number of useful insights. We present them below, numbered, to emphasize the various distinct points.

1. The first implication of these results is that even though red tape is always wasteful, it may be used by the bureaucrat. This is because red tape relaxes the low type's incentive constraint and thereby allows the bureaucrat to charge the low type a higher price.

Red tape in our model is deliberately created by the bureaucrat in order to make money. This contrasts with the view taken by Wilson [1989], among others, who sees red tape as resulting from a set of highly rigid rules set up by the principal in order to limit corruption in the bureaucracy. There is some reason, however, to believe that this cannot be the whole picture. First, in many situations it at least appears that the bureaucrat is going out of his way to generate extra red tape which seems inconsistent with the view that red tape is just a constraint on the bureaucrat. Second, if one takes this view, one still needs to explain why, given that agency problems are ubiquitous, we should not observe the same kind of excessive red tape in private firms as well.¹⁴ By contrast, our view of red tape explains both why bureaucrats favor red tape and why government bureaucracies have more red tape.

While the two views of red tape are very different, it can be argued that they work to reinforce each other. Thus, a rule set up by the principal to limit corruption may be used by a corrupt bureaucrat as an excuse for wasting an applicant's time. To take a concrete and familiar example, most government offices have the rule that anyone who wants anything from the office has to fill out a number of forms. The aim of this rule is to reduce favoritism. Yet the same rule is often invoked by bureaucrats who want to harass certain applicants. They simply ask the applicant to fill out these forms (usually in a large number of copies) and then find small errors in the way the forms were filled out in order to reject the forms so that the applicant has to go through the same procedure again.

14. There is an explanation for this in Wilson [1989], but it relies on the premise that for sociological reasons the government faces certain unique constraints.

2. The second implication of the model is that there would be no red tape if people could pay enough for the slots; i.e., $y \geq L$. In this situation, profit maximization leads to the efficient outcome, and therefore there is no conflict of interest between the bureaucrat and the government. A market failure, then, is necessary for there to be red tape, and of course the same market failure is also the reason why the government is involved in the allocative process.

3. The third implication of the results in the previous section is that, in the world of this model, red tape does not arise because bureaucrats lack incentives. In fact, there is most red tape precisely where the incentives are the strongest; i.e., where F is the largest. This is less paradoxical than it sounds: it is an example of the important observation made in Holmstrom and Milgrom [1991] that increasing the incentives along a dimension of performance that is measurable (here, the share of slots going to the low type) will distort incentives along a nonmeasurable dimension (here, the amount of red tape). In other words, the problem is not that the bureaucrat lacks incentives but that there is a lack of balance between his incentives along different dimensions.

4. A related point is that the most red tape does not arise where the government is the most cynical. If the government were simply interested in making money, it would always set $F = 0$ and allow the bureaucrat to choose the mechanism that maximizes his own income. The government would then recoup the money by charging the bureaucrat a very high price for the slots. We already know that in this scenario there will be no red tape.

This also implies that if the same bureaucracy was a part of a profit-maximizing firm, there would be no red tape.

There will also be no red tape if the bureaucrat shared the government's objective of maximizing social welfare (this is what Claim 1 tells us). It is in the intermediate case, where a welfare-oriented government is trying to control a money-minded bureaucrat, that we would expect to see most red tape. In other words, while a lot of red tape is evidence for some moneymaking by government bureaucrats, it is also evidence that there is some constituency inside the government which is interested in social welfare.¹⁵

15. A referee has pointed out that this result relies on the assumption that a self-serving government has access to a nondistorting mechanism for extracting revenue from the bureaucrat. Absent such a mechanism, even a self-serving gov-

5. High-powered incentives for bureaucrats (high F) in our model lead to better allocations (more H -types get slots) at the cost of higher levels of red tape. In fact, as we remark at the end of the previous subsection, when the cost of red tape to the bureaucrat is small (which seems plausible), even very weak incentives for the bureaucrats can lead to a lot of red tape. This result illustrates a more general point: when goods are being allocated among people who cannot necessarily afford to pay their full value, people will often get goods that are worth more than they have paid for them. As a result, the bureaucrat who is in charge of allocating those goods may be able to make the people who want the goods do something purely wasteful (like enduring some red tape) without reducing what they are willing to pay him. In other words, the bureaucrat has the option of imposing a substantial social cost on his clients at little or no cost to himself. This makes it substantially harder to design proper incentives for the bureaucrat.

6. A consequence of the previous observation is that if the social cost of red tape is sufficiently large, it may be optimal for the government to opt for very low-powered incentives for the bureaucrat. This observation may shed some light on why we do not usually observe explicit high-powered incentives for bureaucrats,¹⁶ and later in the paper (in subsection II.C) we argue that this may be especially true of government bureaucrats in LDCs.

7. Another result follows from equation (1). It is easily checked that T_H is decreasing in y . In other words, red tape will be high where the average person's ability to pay is low. This is because when the ability to pay is low, type H applicants earn very large rents, and therefore a type L applicant is more likely tempted to claim that he is a type H . Therefore, more red tape is needed to discourage him.

Equation (1) also tells us that an increase in N resulting from equiproportional increases in N_H and N_L leads to a rise in red tape. This tells us that red tape will be higher when the slots

ernment may want to set a high value of F just to extract some extra revenue from the bureaucrat. However, while the assumption of a perfectly nondistorting transfer is an idealization, it seems reasonable to assume that since the government and the bureaucrat typically have a long-term relation the transfers between them should be relatively nondistorting even if the bureaucrat is risk-averse and or cash-constrained. As a result, while our results may not hold exactly in a more realistic model, the results from that model should be more or less similar.

16. For other explanations see, for example, Tirole [1992].

are relatively scarcer. This is intuitive: as the slots get scarcer, it becomes more attractive to claim to be a type H (who, as long as $F > 0$, are guaranteed slots).

Both these results hold for any fixed nonzero value of F (when $F = 0$, there is no red tape). The problem is that the assumption of a fixed F is at odds with the structure of the model, since F is actually chosen by the government and typically it will choose different values of F for different levels of the scarcity of the slots and the ability to pay.

The full analysis of the case where F is endogenous is left until subsection II.B. The results we get there are somewhat weaker but along the same lines: the relation between red tape and the ability to pay is still broadly negative, and the relation between red tape and scarcity of the slot is broadly positive.

How do we interpret these relationships? One interpretation is that we are comparing bureaucracies within the same economy who allocate different kinds of goods. Under this interpretation our result for y says that bureaucracies that deal with a population in which the mismatch between the ability to pay and the willingness to pay is the largest¹⁷ will have the most red tape. In particular, this may argue for a lot of red tape in bureaucracies that deal with very poor people.

An alternative interpretation would be to think of low levels of y as representing poorer countries or communities. However, this is not necessarily correct since what matters is the value of y relative to the values of H and L , and while y tends to be lower in poorer countries, H and L may also be lower.

However, as long as we interpret the slots to be beds in a hospital, H and L are naturally interpreted as the value put on life or good health and this, a priori, may be just as high in a poor country as it is in a rich country. If we think of the slots as opportunities for higher education, once again there may not be a tight connection between y and H and L since the latter two numbers are presumably determined, at least in part, in the world market.

There is another reason why y may be low in poorer countries relative to H and L : capital markets work less well in poor coun-

17. This statement is somewhat loose since we do not say how we measure the mismatch. The natural measure is probably the ratio of the two, but this would be strictly correct only if there were no level effects, i.e., if it were true that if we scale down y , L , and H in the same proportion the amount of red tape will be unchanged. However, this is not true for the obvious reason that if the good is not worth very much, no one will be willing to go through much red tape to get it. The interpretation given in the text is therefore less than completely precise.

tries and as a result the ability to pay will tend to be low relative to the willingness to pay.

If we grant the premise that low values of y go with low levels of development, our results suggest a possible explanation of the high correlation, mentioned above, between low levels of development and poor governmental performance.

The interpretation of the results about the effects of an increase in scarcity is more straightforward: bureaucracies that allocate goods that are particularly scarce will be associated with high levels of red tape. In addition, it seems reasonable to think that at least a certain class of publicly provided private good will be scarcer in poorer countries: richer countries will find it easier to expand the supply if there is a perceived scarcity. Thus, in every OECD country every child has access to schooling of a certain minimum quality, but this is palpably not true in LDCs.

8. Finally, the model allows us to give a partial explanation of why government bureaucracies are associated with corruption. As we say in the Introduction, corruption in the government is not inevitable even with self-serving bureaucrats. What causes corruption is the combination of the fact that the bureaucrats want to make money and the fact that governments make laws to prevent them from doing so. It is therefore natural to ask why governments make such laws. One simple answer to this question comes from the model we develop here: red tape in our model results from the fact that the bureaucrats are trying to make money while satisfying the government's imperative of giving every H -type a slot. Therefore, if the government can discourage the bureaucrats from making money by making it illegal to do so, it would also end up controlling the amount of red tape.

Our model thus provides us with a reason why the government would like to impose controls on the prices that the bureaucrat can charge those who want the slots.¹⁸ The model so far does not permit the government to impose such controls, but in Section III we extend the model to allow for them. However, as is reasonable, we do not permit the controls to be perfect, and we put limits on how severely those who breach the controls can be punished. Consequently, unless the controls are essentially nonbinding, some fraction of bureaucrats will charge prices that are above the permitted prices: this is what we call corruption.

We can now investigate the determinants of corruption. In-

18. Holmstrom and Milgrom [1991] make a related argument about why firms may discourage moneymaking by their agents.

tuitively, it would seem that high levels of red tape reflect extreme divergence between the bureaucrat's objectives and what society wants him to do, and therefore it is precisely where red tape is high that we would expect the most corruption. This intuition turns out to be broadly correct, but because of the endogeneity of the government's choice of what kinds of controls to impose on bureaucrats, it is also sometimes possible for red tape and corruption to move in opposite directions.

To the extent that red tape and corruption do move together, our discussion of the determinants of red tape suggests that corruption is most likely in bureaucracies that deal with poor people, in bureaucracies in poor countries, and in bureaucracies that allocate goods that are scarce.

I.E. Plan of the Paper

The exposition of the workings of the model presented in the preceding subsections is misleading in one important respect. We have assumed that the bureaucrat uses winner-pay mechanisms, but because the winner typically values the slot more than he can afford to pay, all-pay rather than winner-pay mechanisms will maximize the bureaucrat's income.

The next section shows that all the results in this section generalize to the case where we allow the bureaucrat to use this broader class of mechanisms. With that assurance at hand, we then return to the case where the bureaucrat only uses winner-pay mechanisms, but we extend the model in other directions. A reader who is impatient about getting to the results may therefore opt to skip Section II on the first reading.

In Section III we look at the case where the government can (imperfectly) observe the payments made to the bureaucrats. This allows us to analyze the determinants of corruption. In Section IV we look at an extension of the basic model where we allow for inequality in the abilities to pay. We conclude in Section V with some discussion of some deficiencies of our model.

II. ANALYSIS OF THE GENERAL MODEL

II.A. Solving the Bureaucrat's Problem

In this section we will solve the bureaucrat's problem assuming that he cares only about his own net income and does not care about social welfare. The other extreme case where the bureau-

crat cares only about social welfare is already addressed in Claim 1.

In solving the bureaucrat's problem, we will take as given the value of the punishment for misallocation, F . By doing so, we can accommodate a range of preferences for the government. For example, in the case where the government itself is money-minded and colludes with the bureaucrat to make money, it would set $F = 0$ so as to not place any additional constraints on the ability of the bureaucrat to make money. On the other hand, by setting F to be very large, the government can essentially force the bureaucrat to allocate a slot to every H -type (though it cannot still control red tape).

The mechanism design problem faced by the bureaucrat is potentially quite complex; however, in a previous version of the paper we show that the optimal mechanism always has a specific form.¹⁹ It can be described by six numbers $(p_H, p_L, \pi_H, \pi_L, T_H, T_L)$ of which the first two represent the price charged to everyone who claims to be a high type or a low type, the second two are the probabilities that a person would get the slot conditional on the person's declared type, and the last pair are the amounts of red tape suffered—once again conditional on the person's declared type.

We can use the fact that each and every slot has to be allocated to eliminate π_L , and as result we can replace π_H by π . With this notation the bureaucrat's maximization problem [MB] can be written as

$$\begin{aligned} & \text{Choose } p_H, p_L, \pi, T_H, T_L \text{ to maximize} \\ & N_H p_H + N_L p_L - N_H v T_H - N_L v T_L - (1 - \pi) N_H F \\ & \text{subject to the constraints} \\ \text{(ICH)} & H \cdot \pi - p_H - \delta T_H \geq H \cdot (1 - \pi N_H) / N_L - p_L - \delta T_L, \\ \text{(ICL)} & L \cdot (1 - \pi N_H) / N_L - p_L - \delta T_L \geq L \cdot \pi - p_H - \delta T_H, \\ \text{(IRH)} & H \cdot \pi - p_H - \delta T_H \geq 0, \\ \text{(IRL)} & L \cdot (1 - \pi N_H) / N_L - p_L - \delta T_L \geq 0, \\ & 0 \leq p_L \leq y, 0 \leq p_H \leq y, 0 \leq \pi \leq 1, T_H, T_L \geq 0. \end{aligned}$$

It is evident from comparing ICH and ICL that, as is common in such incentive problems, these two constraints cannot bind simultaneously as long as the two types are being offered different options. Further, given the fact that the H -type can adopt any

19. Proof is available from the author.

strategy that the L -type has adopted and do strictly better than the L -type, IRH cannot bind. We state this as

LEMMA 1. In any separating equilibrium, ICH and ICL cannot bind simultaneously. IRH never binds in any equilibrium.

The usual analysis of hidden-information models goes on from here to identify the incentive constraint that binds. In our case, however, depending on the values of F and y , either of the incentive constraints may bind. Consider first the case where y is high (higher than L , say). In this case we are in the standard setting where the optimal mechanism is an auction. It both gives the bureaucrat maximal revenue and allocates the slots to the H -types. Therefore, irrespective of the value of F , the chosen mechanism will be an auction, and as is well-known, in the optimal auction the H -type's incentive constraint binds.

The other extreme case is when y is low and F is high. In this setting the bureaucrat's objective is to maximize revenue conditional on every H -type getting a slot. This means that at the optimum the H -types will have a much higher probability of getting the slot than the L -types. If the L -type is to be reconciled to this lower probability of getting the slot, the price he pays must also be significantly lower than the price the H -type pays. Now if y is sufficiently low, the maximal price the L -type can pay is already low, and his participation constraint will not be binding. If this is the case, the bureaucrat will be tempted to raise the price the L -type pays by as much as possible. But there is an obvious tension between this and the need, argued above, to set the L -type's price significantly lower than the H -type's price. As a result, the L -type's incentive constraint will bind in the mechanism chosen by the bureaucrat.

For intermediate values of y and F , either incentive constraint might bind, although from the intuitive discussion in the last paragraph it seems plausible that ICL is more likely to bind when F is high and y is low. Lemma A3 in the Appendix confirms this intuition.

The main analytical goal of this section is to characterize the values of F and y for which there is a high level of red tape. This is complicated by the fact that there are two types of red tape: there is red tape faced by H -types (T_H), and there is red tape faced by L -types (T_L). In principle, depending on which incentive constraint binds, the bureaucrat may want to use either of these

types of red tape (raising T_H relaxes ICL, while raising T_L relaxes ICH). What the next result shows is that the bureaucrat would never want to use red tape against the L -type (the proof is in the Appendix).

CLAIM 2. Self-declared L -types are never subject to any red tape, i.e., there is always an optimum at which $T_L = 0$; and as long as $\nu > 0$, this is the only optimum.

What drives this result is the fact that while more red tape on the L -type does relax ICH, the same effect can be achieved at a lower cost by raising p_L or π . The proof of this result makes use of the fact that the cost of red tape is the same for the two types. Instead, if red tape was much more costly to H -types than it is to L -types, there could be a reason to subject L -types to a little bit of red tape in order to discourage H -types from claiming that they were L -types, and this result would no longer hold.

An obvious consequence of this result is that if red tape is ever used it is used against the H -type. It then follows that if red tape is used at all, it is only used when ICL binds (otherwise there is no reason to use red tape) which happens when F is high and y is low.

To complete the argument, we need to show that when ICL binds the bureaucrat *will* sometimes choose to subject H -types to red tape. This contrasts with the fact that L -types never suffer red tape. The difference between the two cases stems from differences in alternatives to using red tape. In the case of the L -type, the alternative to more red tape was a higher value of π which suits the bureaucrat, since he gets penalized for low values of π . By contrast, in the case of the H -type, the alternative to more red tape was a lower value of π , which hurts the bureaucrat as long as F is positive. As a result, the bureaucrat will be more willing to use red tape.

The final step in the argument is to describe the solution to the bureaucrat's problem. Unfortunately, describing the full solution involves saying what happens in a very large number of different cases. We therefore take the route of describing the full solution in the special case where $\nu = 0$ in the text, while representing the solution to the more general case diagrammatically. The more onerous task of describing the full analytic solution in the more general case is relegated to the Appendix.

CLAIM 3. The solution to the bureaucrat's problem [MB] for the case $\nu = 0$ is as follows:²⁰

- (i) If $y \geq H - (H - L) \cdot (1 - N_H)/N_L$: $\pi = 1$, $p_H = H - (H - L) \cdot (1 - N_H)/N_L$, $p_L = L(1 - N_H)/N_L$, and $T_H = T_L = 0$.
- (ii) If $H - (H - L) \cdot (1 - N_H)/N_L > y \geq L$ and $F \geq L$: $\pi = 1$, $p_H = y$, $p_L = L(1 - N_H)/N_L$, and $T_H = T_L = 0$.
- (iii) If $H - (H - L) \cdot (1 - N_H)/N_L > y \geq L/(N_H + N_L)$ and $0 \leq F < L$: $\pi = [N_L y + (H - L)]/[HN_L + (H - L)N_H]$, $p_H = y$, $p_L = L(H - N_H y)/[HN_L + (H - L)N_H]$, and $T_H = T_L = 0$.
- (iv) If $L > y \geq L \cdot (1 - N_H)/N_L$ and $L \leq F$: $\pi = 1$, $p_H = p_L = y$, and T_H set to solve the equation $L - y - \delta T_H = 0$.
- (v) If $L \cdot (1 - N_H)/N_L \leq y < L \cdot [N_H + N_L]^{-1}$, $L > F \geq 0$: π and T_H set to solve $\pi L - y - \delta T_H = 0$, and $L(1 - N_H \pi)/N_L = y$ and $p_H = p_L = y$.
- (vi) If $L \cdot (1 - N_H)/N_L > y$, for any value of F : the outcome is $\pi = 1$, $p_H = p_L = y$, and T_H satisfying $L - y - \delta T_H = L(1 - N_H)/N_L - y$.

Proof. All the statements except the last one follow from Claims A3 and A4 in the Appendix. The last one requires us to extend the argument slightly, but the extension is sufficiently obvious that we feel that it can be excluded.

The essential features of this solution are as follows: (a) Higher values of F are associated with higher values of π and higher levels of T_H . (b) Higher values of y are associated with lower values of T_H for a fixed F . (c) Higher values of y are not necessarily associated with lower values of π —the highest values of π may obtain at very high and very low values of y . (d) An increase in the scarcity of slots represented by an increase in N_H and N_L in the same proportion, while keeping the number of slots fixed, increases the ratio $N_L/(1 - N_H)$ and thereby increases red tape.

The association between high levels of F and high levels of π is hardly surprising since the point of raising F is to force the bureaucrat to raise π . Higher values of π , ceteris paribus, cause ICL to bind more tightly which then gives the bureaucrat a reason to raise T_H as well. An increase in y allows the bureaucrat to charge higher prices. As a result, he does not need to use as much

20. In writing down this solution, we have implicitly assumed that whenever he is indifferent, the bureaucrat always chooses the socially best outcome.

red tape to induce self-selection by the L -type which is why T_H and y will be negatively associated.

A standard intuition from price theory explains one reason why high values of y result in high values of π ; the high types value the good more, and therefore it pays more to give it to them as long as they can register their preferences as higher prices. When y is low, the reason why the final allocation is very efficient is that it is essentially costless for the bureaucrat to sort the applicants by using red tape.

Scarcity increases red tape because if the slots are scarce, type- L applicants will be more desperate to get the slots. This makes screening harder.

These broad features of the solution to the bureaucrat's maximization problem all turn out to also hold in the more general case, where ν is positive but small relative to δ (this seems to be the natural case to look at). This solution is depicted in Figures I and II, which are based on Claims A3 and A4 in the Appendix.

What changes when ν is large relative to δ ? We show in previous versions of the paper that in this case the outcome is always first-best. This should be intuitive; we have therefore chosen to omit the analysis of this case.

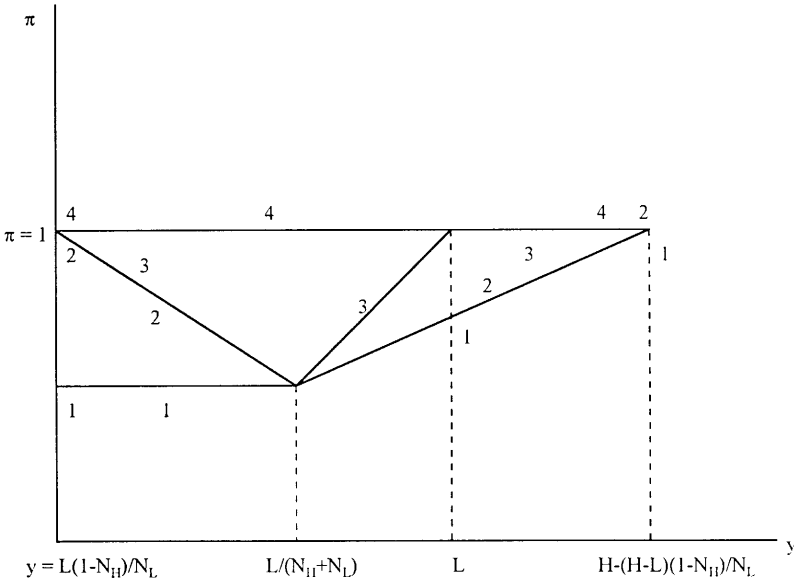
II.B. The Government's Problem

If the government in our model is interested in making money, it will set $F = 0$ and collect the revenue from the bureaucrat as a lump sum fee. When the bureaucrat is welfare-oriented, the choice of F does not matter. The interesting case, therefore, is when the government is welfare-oriented but the bureaucrat is not. The government's maximand in this case will be

$$L + (H - L)N_H\pi(F) - (\delta + \nu)N_H T_H(F),$$

where $\pi(F)$ and $T_H(F)$ are the values of π and T_H that result from the bureaucrat's maximization problem for that particular value of F . In principle, since we have solved the bureaucrat's problem, we can solve the government's problem by comparing the government's maximand for different values of F . In practice, this will involve considering a very large number of cases. We therefore only look at the government's problem in the special case where $\nu = 0$, which makes the problem much more tractable.

It is evident from Claim 3 that in this case the government need only choose between $F = 0$ and $F = L$. Furthermore, for



Curve 1: $F < L(v/\delta + vN_H/\delta N_L)$

Curve 2: $L(v/\delta + vN_H/\delta N_L) \leq F < L$

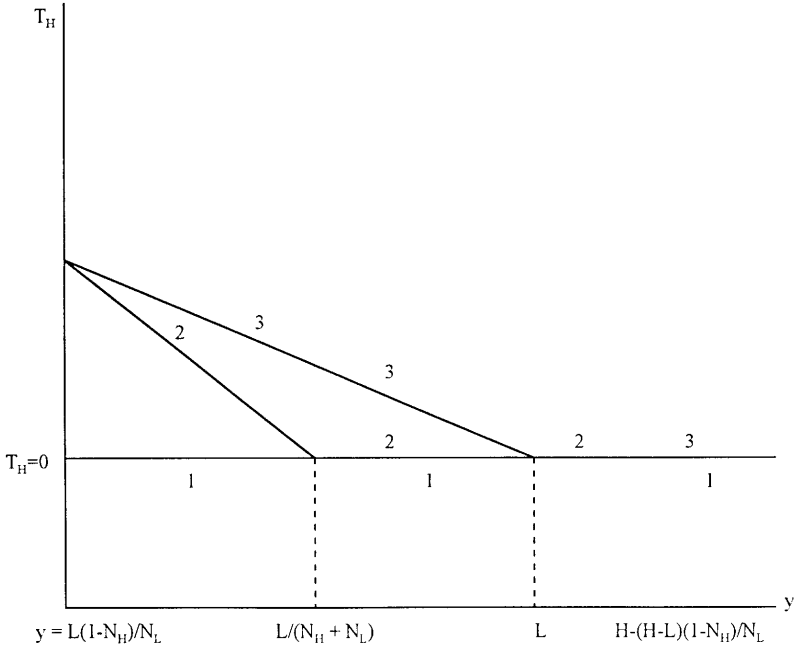
Curve 3: $L \leq F < L(1 + v/\delta)$

Curve 4: $L(1 + v/\delta) \leq F$

FIGURE I
 π as a Function of y

extreme values of y , i.e., $y \geq H - (H - L) \cdot (1 - N_H)/N_L$ and $y < L(1 - N_H)/N_L$, the value of F does not matter—all values of F result in the same outcome. In both these cases the government will presumably choose $F = 0$; i.e., let the bureaucrat do whatever he wants.

For values of y between $H - (H - L) \cdot (1 - N_H)/N_L$ and L , the solution is also straightforward. It is evident from the comparison of cases (ii) and (iii) in Claim 3 that in this case a higher value of



Curve 1: $F < L(v/\delta + vN_H/\delta N_L)$

Curve 2: $L(v/\delta + vN_H/\delta N_L) \leq F < L(1 + v/\delta)$

Curve 3: $L(1 + v/\delta) \leq F$

FIGURE II
 T_H as a Function of y

F is always preferable since it generates a higher value of π without generating any red tape.

The less obvious case is when y is between L and $L(1 - N_H)/N_L$. In this case a simple calculation based on a direct computation of the government's maximand for the two values of F establishes that $H > 2L$ is a sufficient condition for always using $F = L$. However, if $H < 2L$, $F = 0$ will be used as long as y is between $L(1 - N_H)/N_L$ and $L \cdot [N_H + N_L]^{-1}$, but for higher values of y , $F = L$ is still optimal. The chosen value of F is always weakly increas-

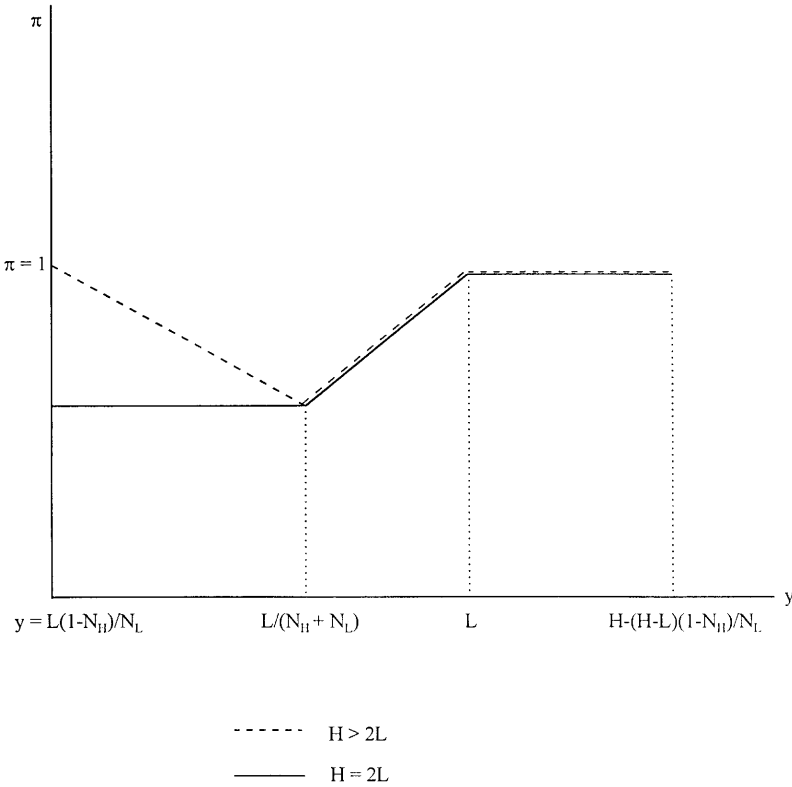


FIGURE III
 π as a Function of y When F Is Endogenous

ing in H keeping L fixed: this is because a high H makes it more important for each person of type H to get a slot.

How does the relation between π , T_H , and y look now that F is endogenous and depends on y ? These are given in Figures III and IV for two cases: $H > 2L$ and $H = 2L$ (with the interpretation that $H = 2L$ is the limit of the case where $H < 2L$ and represents all such cases). It should be evident from the discussion above that these are essentially the two canonical cases. In the case where $H > 2L$, the pictures are exactly the same as they were when F was exogenously set to be greater than or equal to L . However, in the case where $H = 2L$, endogenizing F does change the picture since, at low levels of y , $F = 0$ is chosen but at higher values the chosen value of F goes up to L . As a result, an increase in y over a certain range causes T_H to go up.

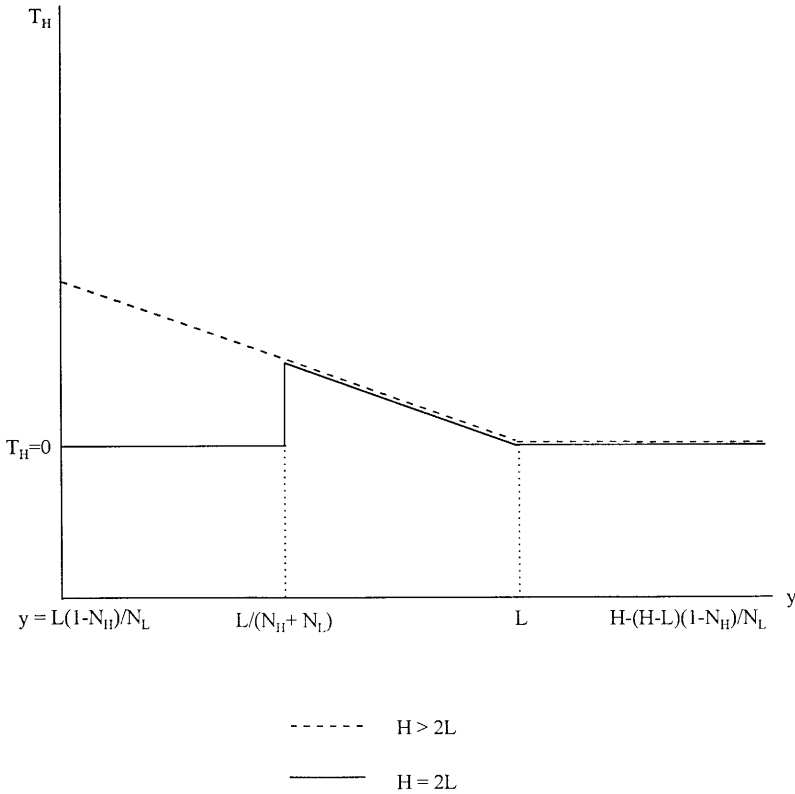


FIGURE IV
 T_H as a Function of y When F Is Endogenous

We have not explicitly considered the effect of changes in the scarcity of the slots, but it can be shown that the effect of an increase in the scarcity of the slots is similar to that of a fall in y . It typically leads to a rise in the level of red tape, but it may also cause F to fall, and as a result, for a specific range of parameter values, red tape may be lower even though the slots are scarce.

II.C. What Do We Learn from the Results of the More General Model?

The results here largely confirm what we found in the simpler version of the model analyzed in subsection I.C. As before, for a fixed value of y , an increase in F leads to a higher level of red tape. Combined with Claim 1, this confirms our earlier claim that red tape is maximized when there is a conflict of objectives

between the government and the bureaucrat (with the government being welfare-oriented and the bureaucrat self-serving). It also confirms that there would be no red tape if, instead of the government, a private firm were carrying out the allocation (a private firm would set $F = 0$). Of course, the overall outcome would be worse.

However, note that the effect of an increase in F on the level of red tape in this model is much less dramatic than it was in the model in subsection I.C. There, for $\nu = 0$, any positive value of F leads the bureaucrat to go immediately to the maximum level of red tape that he would ever use for that level of y . Here, as is evident from Figure II and Claim 3, the response is more gradual. This is because the use of all-pay rather than winner-pay mechanisms allows the bureaucrat to extract more of the surplus from the applicants, which then makes the bureaucrat internalize more of the cost of the red tape he imposes on them. This suggests that a movement toward creating an environment where bureaucrats can use all-pay mechanisms to allocate scarce publicly provided private goods may actually help improve bureaucratic performance.

As in subsection I.C for a fixed value of F , there is a negative relation between y and red tape. The analysis in this section goes beyond the previous analysis in endogenizing F . Endogenizing F does not change the relation between y and red tape as long as H is sufficiently greater than L . However, when H is close to L , the relation between red tape and y may be nonmonotonic, although it will still continue to be true that very low values of y will be associated with very high levels of red tape and red tape will be absent at high levels of y . The relation between red tape and the scarcity of slots is similar to that between red tape and y , with low levels of y corresponding to high levels of scarcity.

The behavior of π as a function of y can be read from Figure I and turns out to be subtler than one would have predicted from the preliminary analysis: except when F is very high (when $\pi = 1$ at all values of y in our range) or very low (when π is constant at low levels of y), π is always U-shaped as a function of y ; it is high at high values of y as well as at low values of y and is lower in between. The relationships are more or less the same with F endogenized (see Figure III).

Since we allow the government to choose F , we also find conditions under which a welfare-oriented government will deliberately choose low-powered incentives for the bureaucrat (i.e., set

$F = 0$) in order to avoid generating too much red tape. This will happen when the difference between H and L is not too large (i.e., the misallocation is not too costly), the slots are scarce, and y is relatively small (which imply that the bureaucrat, if pushed, will use high levels of red tape).

Since we have taken the view that lower values of y and greater scarcity go with lower levels of development, this result suggests that at least in situations where the cost of misallocation is small, bureaucrats in less developed countries will tend to have weaker incentives than their counterparts in the developed world. We can also read the model as saying that within the same country, those bureaucracies that deal most with people with low abilities to pay (relative to their willingness to pay), will have the weakest incentives.

III. TOWARD A THEORY OF CORRUPTION

To restore efficiency in the economy modeled here, the government will need to be able to control the prices charged by the bureaucrats. We will now modify our model to allow the government some possibility of observing the payments that are made to the bureaucrats.

We introduce the possibility of monitoring payments to bureaucrats by assuming that with some probability $\phi < 1$, the government finds out about *the mechanism being used by the bureaucrat to allocate the slots* (here we are using the word mechanism in its broader sense so that if the bureaucrat uses several different rules to allocate to different people, we will consider them together to be a part of a single mechanism). Recall that we have already assumed and continue to assume that the government knows the fraction of type L applicants who got a slot. What knowing the mechanism tells the government is whether the bureaucrat is charging the recommended prices or whether he is asking for additional bribes.²¹

In this setting, if the government could also inflict arbitrarily large punishments on the bureaucrats, it is easy to see that it could always implement the optimal outcome. All it would have to do is to recommend that the bureaucrat uses the optimal mech-

21. It also tells the government how much red tape is being used, but this is not extra information, since once it knows the prices and the allocation, it can always infer the amount of red tape.

anism and to punish any detected deviation from this mechanism with such severity that no bureaucrat would ever contemplate deviating.

The more interesting case is the one where there is a bound on how much a bureaucrat can be punished. We model this by assuming that there is an institutionally given worst punishment that the government can inflict on any bureaucrat (this may be the loss of his job and a prison stay of several years). Denote the utility level of a bureaucrat who is undergoing this punishment by B , and assume that there is a distribution function $G(B)$, which gives the fraction of the population of bureaucrats whose lower bound is no higher than B .²² We will assume that B is private information.

There are a number of alternative patterns that can emerge in this setting and investigating all of them is beyond the scope of this paper. Here we confine ourselves to the situation where the government wants to allocate the slots efficiently even at the cost of some red tape (this is the case where H is large relative to L).

To simplify the analysis further, let us revert to the assumption made in subsection I.C limiting the bureaucrat to winner-pay mechanisms. Also, to limit the number of cases, assume that $L \geq y \geq L - L(1 - N_H)/N_L$ and $v = 0$.

Under these assumptions, all mechanisms that achieve the efficient allocation of slots take the form $\{p_H, p_L, T_H\}$, where p_H and T_H are the price and red tape assigned to a type H applicant (who always gets a slot) and p_L , the price paid by a type L , satisfies the incentive compatibility constraint;

$$L - p_H - \delta T_H = (L - p_L)(1 - N_H)/N_L.$$

Of these mechanisms, the one that is least likely to lead to corruption is the one that sets the highest prices for both types (the higher the official price, the less people will want to pay in excess of that price to increase their chances of getting the slot). Therefore, p_H should be set equal to y .

Now suppose that the government announces a mechanism $\{y, p_L^*, T_H^*\}$. In other words, it sets both the prices the bureaucrat is allowed to charge and the maximum amount of red tape that the bureaucrat is permitted to use (an example of a government rule about how much red tape is permitted is the rule recently

22. Those with low levels of B may be thought of as those who especially value their reputation for being honest.

introduced in India requiring all passport applications to be processed within a certain number of days). We assume that the mechanism recommended by the government is incentive compatible from the point of view of the applicants. Once such a mechanism is announced, bureaucrats are required by the government to implement that mechanism, and it is also announced that any bureaucrat who is caught deviating from this mechanism will receive the maximal punishment.²³

The government also needs to choose F . In deciding on F , the government can take advantage of the fact that if $\nu = 0$ and the bureaucrat only uses winner-pay mechanisms, the bureaucrat will always give every type H applicant a slot for any strictly positive value of F . This was shown in subsection I.C and continues to hold in our current model. Moreover, it holds *irrespective of whether the bureaucrat follows the mechanism the government wants him to follow*: the only difference is that if he chooses to deviate, he will use red tape to screen out L -type applicants instead of relying on prices.

Given the assumption, made above, that H is large relative to L , the government will always set a nonzero level of F . Given that it is indifferent between all nonzero levels of F , assume now that it sets the value of F to be so close to 0 that the expected value of the fines can be ignored while calculating the bureaucrat's utility level (consequently, we do not need to worry how the bureaucrat can be fined in the state of the world where he is already being punished for taking bribes).

Given all these assumptions, the bureaucrat who will be on the margin of deviating and asking for a bribe, will have a B which satisfies

$$N_H y + (1 - N_H) p_L^* = (1 - \phi) y + \phi B.$$

Clearly, the left-hand side of this equation represents the utility of a bureaucrat who follows the rules while the right-hand side represents the utility of a bureaucrat who, instead, charges y for

23. The mechanism used here is an efficiency wage-type mechanism first used in the context of corruption by Becker and Stigler [1974]. It is in principle possible to allow the government to use more sophisticated mechanisms (such as a linear or nonlinear tax on the bureaucrat's income from selling the slots) which may actually work better. We justify not using such mechanisms on the grounds that we do not observe such mechanisms (we also believe that so long as the government has limited ability to observe the bureaucrat's income, the results will not change very much even if we change the model in this direction). For a more detailed discussion of the kinds of incentive schemes used by governments vis-à-vis their bureaucrats, see Banerjee [1995], Kofman and Lawarree [1990], and Tirole [1992].

every slot and gets caught with probability ϕ . Solving for the value of p_L^* using the incentive-compatibility constraint, gives us

$$p_L^* = L - N_L[L - y - \delta T_H^*]/(1 - N_H).$$

Substituting this expression into the above equation gives us

$$N_H y + (1 - N_H)L - N_L[L - y - \delta T_H^*] = (1 - \phi)y + \phi B,$$

which can be written in the form

$$(2) \quad (N - 1)(L - y) - \delta N_L T_H^* = \phi(y - B).$$

Denote the value of B that solves this equation by B^* . Clearly, those and only those with values of B greater than this critical value will choose to break the rules and ask for bribes. In other words, $1 - G(B^*)$ measures the extent of corruption in this economy.

Note that the corruption that arises here is in a very direct sense created by the government. The government creates corruption by imposing a rule on the bureaucrats that some bureaucrats will follow and others disregard: if there were no such rule, there would be no bribes and no corruption. Nevertheless, the reason why the government chooses to impose this rule is that it helps it fight wasteful red tape in the bureaucracy.

This contrasts with the quite common view that corruption arises, at least in part, out of a need to get around the red tape that is endemic in government bureaucracies.²⁴ In this view, what causes red tape is something that is usually exogenous and explained, if at all, by reference to the sociology of the government. There is therefore little one can do about red tape itself, and anything that helps get around it is probably a good thing. Fighting corruption, in this view, may therefore be a bad thing.

By contrast, our view is that a lot of red tape is deliberately created by the bureaucrats in order to make more money. Fighting corruption, by limiting the amount of money the bureaucrat can make, may therefore also reduce red tape.

A second implication of this analysis of corruption is that corruption only arises when the government has a reason to try to limit moneymaking by bureaucrats. In our model, if the government was indifferent to social welfare and only interested in

24. For a forthright if somewhat dated statement of this view, see Nye [1979] or Leff [1979]. See Waterbury [1979] for a critique of this view on empirical grounds.

making money, there would be no corruption. Like red tape, corruption arises from a conflict of interest.

A number of other conclusions follow from equation (2). First, B^* is increasing in y for any fixed value of T_H^* and the other parameters. In other words, everything else remaining the same, a fall in y increases corruption. In other words, somewhat paradoxically, there is more illegal moneymaking precisely when there is less money around. This is because an increase in y enables the government to raise the legal price paid by a type L applicant by more than the original increase in y (see the expression for p_L^* given above).

Second, a simple calculation establishes that an equiproportional increase in N_H and N_L , for any fixed value of T_H^* and the other parameters, reduces B^* and therefore increases corruption. In other words, there is more bribery as the good being allocated becomes scarcer.

Third, once again keeping T_H^* fixed, an increase in y or an equiproportional fall in N_H and N_L will lead to a fall in the total amount of red tape. To see this, observe that the average amount of red tape suffered by an H -type applicant is given by

$$(3) \quad G(B^*)T_H^* + (1 - G(B^*))T_H.$$

The first term in this expression is the amount of red tape that is associated with bureaucrats who do not deviate from the recommended mechanism, and the second term comes from those who do deviate. Now, both the increase in y and the fall in N_H and N_L have the effect of reducing the fraction of those who take bribes and therefore lead to a fall in red tape (because $T_H \geq T_H^*$). Also, as shown in subsection I.C, both these changes have the effect of reducing T_H , which goes in the same direction.

However, the above results about the effect of a fall in y , or an increase in N_H and N_L assume that the permitted amount of red tape, T_H^* , is exogenously fixed. This is misleading since in our model the government chooses T_H^* and an increase in T_H^* by itself, increases B^* and therefore reduces bribery.²⁵ We therefore need to treat T_H^* as an endogenous variable when we do the comparative statics. Since, in the situation considered in this section, all bureaucrats (whether or not they take bribes) allocate the slots in the same way, the government, in choosing T_H^* , needs only to look at the effect on the average amount of red tape. Differentiat-

25. This is because an increase in T_H^* allows p_L^* to be increased.

ing the expression given in equation (3) for the average amount of red tape, with respect to T_H^* , yields the first-order condition:

$$(4) \quad G(B^*)/G'(B^*) = (T_H - T_H^*)\delta N_L/\phi.$$

Equation (2) embodies a very simple trade-off: an increase in T_H^* hurts those already dealing with uncorrupt bureaucrats, but it also increases the fraction of bureaucrats who are not corrupt. Therefore, as the equation makes evident, what matters for the choice of T_H^* is the population of inframarginal (uncorrupt) bureaucrats relative to the population of those who are at the margin of becoming uncorrupt. T_H^* will tend to be high when there are lots of marginal bureaucrats relative to the number of those who are inframarginal.

The effect of an increase in y on T_H^* turns out to be impossible to sign on purely a priori grounds because, while an increase in y increases B^* and therefore increases the number of inframarginal bureaucrats, it also affects the number who are at the margin and the net effect on $G(B^*)/G'(B^*)$ is ambiguous. However, for a large range of distribution functions, $G(\cdot)$ (including, for example, the case where the underlying density is uniform), it can be shown that T_H^* falls when y goes up. Furthermore, it is possible to construct examples where the fall in T_H^* resulting from the increase in y is so large that it swamps the direct effect of the increase y on B^* and the net effect on B^* is negative. In other words, an increase in y can lead to an *increase* in corruption because of the endogeneity of T_H^* . For exactly the same reasons, a fall in the scarcity of the good can actually lead to an increase in corruption.

These kinds of “perverse” comparative statics results are less likely to arise if the density function corresponding to the $G(\cdot)$ function has a mass point (or a highly concentrated density) at the lowest point in its support but nowhere else. This kind of density captures the plausible idea that the population of bureaucrats contains a hard core of incorruptible people, but otherwise there is a lot of diversity in how people feel about getting caught taking a bribe. In this case there will always be a large number of inframarginal bureaucrats, and therefore it is costly to raise T_H^* in order to combat corruption. As a result, it is unlikely that when y falls, T_H^* will be raised by so much that there will actually be a fall in corruption.

It is also worth remarking that even with F endogenous,

there will be no corruption in the case where y is higher than L since in this case there is no conflict between making money and furthering social welfare. Thus, the negative relation between y and corruption holds at least when we compare very high and very low levels of y .

As we noted above, the direct effect of an increase in y on red tape is always negative. In addition, we just argued that an increase in y typically leads to a fall in T_H^* which reinforces this effect. However, in the scenario where an increase in y increases corruption, this increase in corruption can increase red tape. However, note that this effect needs to be strong enough to swamp the other two effects if the overall effect of an increase in y is to increase red tape. This seems somewhat implausible.

To summarize, once we endogenize the permitted amount of red tape, we no longer get the simple unambiguous comparative statics results that we got when the permitted amount of red tape was taken as exogenously given. The amount of corruption and somewhat less plausibly, the amount of red tape, may actually go up when the applicants have a higher ability to pay or the slots are less scarce. This is because the government responds to the increase in the ability to pay or the fall in scarcity by severely limiting the amount of red tape the bureaucrat is allowed to use. In a sense, what is going on is that the bureaucrats' effective incentive scheme is becoming much more demanding, and this leads to an outcome where more bureaucrats fail to meet the standard.

We also identify one quite reasonable setting where an increase in the ability to pay or fall in the scarcity of the slots always reduces both corruption and red tape. This is the situation where the population of bureaucrats contains a core of people who are completely incorruptible.

IV. IMPLICATIONS OF INEQUALITY IN THE ABILITY TO PAY

We have so far ignored the possibility that different people may have different abilities to pay. This is an important deficiency since a standard justification of red tape-like procedures is that they protect the poor.²⁶ The conclusions of this section are as follows. (i) The presence of inequality increases the amount of red

26. See, for example, Weitzman [1977].

tape used by both a profit-minded government and the government in our model, and (ii) it remains true that more red tape is used when the government is welfare-oriented.

There are at least two ways to introduce inequality into this model. The simpler case is where both the bureaucrat and the government can observe each applicant's ability to pay. In this case the government sets an F that depends on the applicant's ability to pay, and the bureaucrat chooses a different mechanism depending on the applicant's ability to pay. The bureaucrat's problem then consists of a number of parallel problems of the type we solve in the previous section. It is easy to see that the outcome of the bureaucrat's maximization problem will be such that those who have less money (smaller y) will face more red tape.

This conclusion gets reinforced if we assume that neither the government nor the bureaucrat can observe the applicant's ability to pay. Assume that the ability to pay takes two values, y_1 and y_2 ($y_1 > y_2$) with probabilities μ and $1 - \mu$, and that a person's valuation of the slot is statistically independent of his ability to pay. Also to make the problem interesting, assume that $1 > \mu(N_H + N_L)$; i.e., there are not enough rich people to fill up the slots (if we do not make this assumption, the poorer people may be irrelevant). In all other respects let the model be exactly the same as the model we introduce in Section I (in other words, we do not allow the government to observe payments to bureaucrats so that the question of corruption does not arise).

This is a two-dimensional screening problem, and these are notoriously difficult to solve. To make it tractable, we make the simplifying assumption we made in the introduction, namely, that the bureaucrat is limited to winner-pay mechanisms. We also assume that $\nu = 0$ and that $y_1 < L$.

With these simplifying assumptions the problem turns out to be quite simple to solve. Given that we assume that $y_1 < L$ and that only those who get the slot pay for it, the individual rationality constraints will not bind for any of the agents. Therefore, the bureaucrat can impose some extra red tape on the agents without having to cut the price he charges them. Since in addition we have assumed that $\nu = 0$, extra red tape also costs the bureaucrat nothing. Therefore, a self-serving bureaucrat will always charge the applicants the highest price they can pay and then use red tape to ensure that the mechanism he sets up is incentive compatible.

The problem faced by a profit-minded government with a profit-minded bureaucrat therefore has a simple solution; the bureaucrat will set two prices, y_1 and y_2 (i.e., the maximum possible prices), and offer a slot to each person who pays the higher price and randomly select $1 - \mu(N_H + N_L)$ persons among those who offer to pay the lower price. This will be incentive compatible if²⁷

$$(2) \quad L - y_1 \geq L[(1 - \mu(N_H + N_L))/(N_H + N_L)] - y_2.$$

If not, the bureaucrat will have to threaten those who pay less with some red tape; the exact amount of red tape, T , will be given by

$$(3) \quad L - y_1 = L[(1 - \mu(N_H + N_L))/(N_H + N_L)] - y_2 - \delta T.$$

In the conflicting objectives model, if the government sets a high enough F , the bureaucrat will want to give a slot to every high type. The mechanism that maximizes the bureaucrat's profits conditional on giving a slot to every high type, will be described by four triplets $(y_1, T_1, 1)$, $(y_1, 0, \min\{0, (1 - N_H)/\mu N_L\})$, $(y_2, T_2, 1)$ and $(y_2, 0, \min\{(1 - N_H - \mu N_L)/(1 - \mu)N_L, 0\})$ with T_1 and T_2 satisfying

$$(4) \quad L - y_1 - \delta T_1 = (L - y_1)[\min\{(1 - N_H)/\mu N_L, 1\}]$$

$$(5) \quad L - y_2 - \delta T_1 = (L - y_2)[\min\{(1 - N_H - \mu N_L)/(1 - \mu)N_L, 0\}].$$

The first number of each of these triplets is the price that a person who chooses that option pays. The second number is the amount of red tape he has to go through. The last number is the probability he gets a slot. The first triplet is what a rich high type chooses, the second what a rich low type chooses, the third is what a poor high type chooses, etc. Note that each type is paying the maximum amount he can pay.

The outcome generated by this mechanism is that the rich high types and the poor high types all get slots. If the number of remaining slots is less than the number of rich low types, we assume that the rich low types get all of these slots. If there are slots left over after all the rich low types have chosen, then they will be given to some of the poor low types.

The outcome in the case where both the government and the bureaucrat are welfare-oriented is still going to be socially efficient as long as y_2 satisfies $y_2 \geq L - L \cdot (1 - N_H)/N_L$ since in this

27. It is easily checked that this is the incentive constraint that may bind.

case we can use the mechanism used in the argument for Claim 1, with y_2 substituted for y .

This quite rudimentary analysis yields a number of useful insights.

1. A comparison of equations (4) and (5) with equation (3) establishes that while in the presence of inequality red tape will arise in both the self-serving government model and the conflicting objectives model, there will always be more red tape generated under the latter model. This confirms the results in the previous sections.

2. It is evident from equations (4) and (5) that an increase in inequality in the distribution of y , keeping the mean unchanged, reduces T_1 and increases T_2 , but on balance, the social waste due to red tape always goes up. This is shown in the Appendix (See Claim A5). The reason is that the probability that a poor low type gets a slot is lower than the probability that a rich low type gets the slot. As a result, a poor low type has more of an incentive to claim that he is a high type than the rich low type. Moreover, and for the same reason, a change in y has a bigger impact on the poor low type's incentive to misrepresent his type than it has on the corresponding incentive for the rich low type. As a result, the change in the red tape for the poor low type will also have to be larger than the corresponding change for the rich low type. Consequently, the rise in red tape caused by the fall in the poor low types' ability to pay will dominate the fall in red tape resulting from the rise in the rich low types' ability to pay.

3. The poor face more red tape than the rich in the conflicting objectives model. The same result may also be true in the pure self-serving government model, but only if y_2 is sufficiently low. In both cases the bureaucrat uses this extra red tape to threaten the rich, so that the rich are forced to buy their way out of it.

4. The poor of the low type get less access to the slots than the rich of the low type both under the conflicting objectives model and the pure self-serving government model, although the difference in access is greater in the latter case.

V. CONCLUSIONS

The model proposed in this paper, while both simple and stylized, makes a number of predictions that broadly fit the pattern of what we know about misgovernance. However, it also has a number of important and obvious limitations.

An implication of this model is that governments in developed countries should use the market more than in LDCs. While this is true in some cases,²⁸ there are others like health-care where the market is not used. Of course, our model only tells us the efficient outcome and ignores distributional considerations that may explain why the market is not used. It is still a puzzle why, given that the market is not used, there is so little corruption in the health-care bureaucracy in most OECD countries. The explanation suggested by our model is that there is an adequate supply of health-care, i.e., the good is not scarce enough to make corruption worthwhile. Whether this is the right story is an open empirical question.

Our model also does not deal with the issue of whether there are cultural or institutional determinants of government performance. One stereotype we did not take up (because it concerns preferences rather than outcomes) is the characterization of third world societies as being much more casual about corruption in government than first-world governments. It has been pointed out that in this instance what appears to be cultural and exogenous may be endogenous and rational in the sense that there may be multiple equilibria in some of which corruption may be rare and heavily punished and others in which corruption is common and tolerated.²⁹

Of course, even if we accept the multiple equilibrium view, it remains to explain why the culture of corruption should emerge principally in LDCs.³⁰ Two explanations come to mind: one could argue that the culture of corruption is what causes LDCs to be less developed. This we find somewhat implausible given that these LDCs also tended to be poor countries before the recent era of large-scale government interventions in the economy. The other, more convincing (to us) theory holds that development is a process of transforming a large complex of institutions along with increasing the GNP. The culture of corruption in poor countries is at least partly a result of underdeveloped institutions (like a lack of democracy).

28. Few rich countries have licenses for production and imports; and in the United States, for example, oil drilling rights are auctioned off too.

29. See Cadot [1987], Clague [1993], and Sah [1991] for different arguments within this broad category. Tirole [1996] provides a model within which a temporary increase in corruption may become irreversible. Also see Acemoglu [1992] and Murphy, Shleifer, and Vishny [1993] for the related argument that the presence of corruption may actually induce others to become corrupt by reducing the return to the honest activity.

30. Japan and Italy being well-known exceptions.

APPENDIX

Proof of Claim 2

The only interesting case is the one where there is a separating equilibrium. There is no reason to use red tape in a pooling equilibrium.

Now note that if ICH does not bind, then the bureaucrat will always want the value of T_L to be lower. Therefore, $T_L > 0$ implies that ICH binds which in turn implies that ICL does not bind so that $T_H = 0$.

Next observe that if IRL does not bind, we must have $\pi = 1$ because, if not, it is always possible to raise π and relax all the binding constraints. It is also easy to see that if IRL does not bind, we must have $p_L = y$ since otherwise it would be possible to raise p_L and relax all the binding constraints while making the bureaucrat better off.

Consider first the case where IRL does not bind so that $\pi = 1$ and $p_L = y$. Then $H\pi - p_H = H - p_H > H \cdot (1 - N_H)/N_L - y - \delta T_L$ so that ICH does not bind.

Next consider the case where IRL binds. For the reason given in the previous paragraph, we cannot have $\pi = 1$ and $p_L = y$. First, consider the option $p_L < y$. Then an increase in p_L , combined with a reduction in T_L keeping $p_L + \delta T_L$ constant, always improves the outcome.

Finally, consider the possibility that at the optimum $\pi < 1$. In this case increase π while reducing T_L so as to keep the IRL binding. Then $d\pi/dT_L$ will satisfy $(L \cdot N_H/N_L)d\pi/dT_L = -\delta$. Substituting this into the ICH constraint, we find that the left-hand side goes up (because π goes up) and the right-hand side goes down. Therefore, this change relaxes the ICH constraint, and it is always optimal to make such a change. This proves the first part of our claim. The second part follows from the fact that with $\nu > 0$ a reduction in T_L is strictly in the bureaucrat's interest.

Proved

Solving the Bureaucrat's Maximization Problem [MB]

We solve the bureaucrat's maximization problem [MB] in a number of steps. The first step in solving the bureaucrat's maximization problem is to consider the more limited maximization problem where we drop the constraint ICL. This gives us the problem [mb]:

Choose p_H, p_L, π, T_H, T_L to maximize
 $N_H p_H + N_L p_L - N_H v T_H - N_L v T_L - (1 - \pi) N_H F$,
 subject to the constraints

- (ICH) $H \cdot \pi - p_H - \delta T_H \geq H \cdot (1 - \pi N_H) / N_L - p_L - \delta T_L$,
- (IRH) $H \cdot \pi - p_H - \delta T_H \geq 0$,
- (IRL) $L \cdot (1 - \pi N_H) / N_L - p_L - \delta T_L \geq 0$,
- $0 \leq p_L \leq y, 0 \leq p_H \leq y, 0 \leq \pi \leq 1, T_H, T_L \geq 0$.

The solution to this problem is given in Claim A1.

CLAIM A1. The solution to the problem [mb] described above is given below.

If $F \geq L$ and $y \geq H - (H - L) \cdot (1 - N_H) / N_L$, $\pi = 1$, $p_H = H - (H - L) \cdot (1 - N_H) / N_L$, $p_L = L(1 - N_H) / N_L$, and $T_H = T_L = 0$.

If $F \geq L$ and $H - (H - L) \cdot (1 - N_H) / N_L > y > L(1 - N_H) / N_L$, $\pi = 1$, $p_H = y$, $p_L = L(1 - N_H) / N_L$, and $T_H = T_L = 0$.

If $F \geq L$ and $y \leq L(1 - N_H) / N_L$, $\pi = 1$, $p_H = y$, $p_L = y$, and $T_H = T_L = 0$.

If $F < L$ and $y \geq H - (H - L) \cdot (1 - N_H) / N$, $\pi = 1$, $p_H = H - (H - L) \cdot (1 - N_H) / N_L$, $p_L = L(1 - N_H) / N_L$, and $T_H = T_L = 0$.

If $F < L$ and $H - (H - L) \cdot (1 - N_H) / N_L > y \geq L / (N_H + N_L)$, $\pi = [N_L y + (H - L)] / [H N_L + (H - L) N_H]$, $p_H = y$, $p_L = L(H - N_H y) / [H N_L + (H - L) N_H]$, and $T_H = T_L = 0$.

If $F < L$ and $L / (N_H + N_L) > y \geq L(1 - N_H) / N_L$, $\pi = (1 - N_L y / L) / N_H$, $p_H = y$, $p_L = y$, and $T_H = T_L = 0$.

If $F < L$ and $y < (1 - N_H) / N_L$, $\pi = 1$, $p_H = y$, $p_L = y$, and $T_H = T_L = 0$.

Proof of Claim A1

Observe that at the optimum either the IRL constraint binds or $p_L = y$ (otherwise the bureaucrat would raise p_L). Consider first the case where the IRL constraint binds at the optimum. Assume to start out that the ICH constraint does not bind. Then p_H must be equal to y . What remains to be determined is the value of π . If ICH is not binding, a reduction in π has two effects: it increases $p_L N_L$ by $L \cdot N_H$, and it increases the expected punishment term by $F \cdot N_H$. Therefore, if $L \leq F$, π will be set equal to 1. If $L > F$, π will be reduced until either ICH binds or IRL stops binding so that it ceases to be profitable to reduce π .

This leaves us with four distinct cases we need to consider:

- i) $F \geq L$, and IRL binds;
- ii) $F \geq L$, and IRL does not bind;
- iii) $F < L$, and IRL binds;
- iv) $F < L$, and IRL does not bind.

Consider the first two cases together. We know from above that if $F > L$ and IRL binds, π will be set equal to 1; a fortiori this will also be true if IRL does not bind. Then if IRL were to bind, p_L would be $L(1 - N_H)/N_L$. Therefore, IRL binds if and only if $L(1 - N_H)/N_L \leq y$.

Let IRL bind: then from ICH, $H - p_H \geq (H - L) \cdot (1 - N_H)/N_L$ which implies that $p_H \leq H - (H - L) \cdot (1 - N_H)/N_L$. Now either this is an equality or $p_H = y$. Which happens depends on how y compares with $H - (H - L) \cdot (1 - \pi N_H)/N_L$; p_H will be the smaller of the two.

If IRL does not bind, then $p_L = y$. Then ICH cannot bind either since $H(1 - N_H)/N_L - y < H - p_H$. Therefore, $p_H = y$.

Turning now to the case where $F < L$ and both ICH and IRL bind, we substitute IRL in ICH to get

$$(A1) \quad H \cdot \pi - p_H = (H - L) \cdot (1 - \pi N_H)/N_L.$$

If we increase p_H toward y , π has to go up. The rate at which it goes up, $d\pi/dp_H$, is $1/[H + (H - L)N_H/N_L]$. The resulting reduction in p_L will be $L \cdot (N_H/N_L) \cdot [H + (H - L)N_H/N_L]^{-1}$. Therefore, there will be a net gain from the increase in p_H if $N_H > N_L \cdot L \cdot (N_H/N_L) \cdot [H + (H - L)N_H/N_L]^{-1}$ which is always true. So, the outcome in this case is either $p_H = y$ or $\pi = 1$.

Which of these two outcomes obtains at the optimum depends on which binds first as we increase p_H toward y . It can be checked by looking at (A1) that if y is greater than $H - (H - L) \cdot (1 - N_H)/N_L$ then π will hit 1 before p_H hits y . Therefore, this will be the outcome. However, if y is below this critical level, then p_H will hit y with π less than 1.

Of course, these predictions assume that the IRL constraint binds rather than the alternative outcome $p_L = y$. Now so long as y is greater than L , we cannot have $p_L = y$ since this would violate IRL. Therefore, the IRL constraint must bind if y is higher than L . By continuity it will also continue to bind when y is lower than L but not too low. However, as we continue to reduce y , π will fall toward $(1 - \pi N_H)/N_L$, and p_L will rise to close the gap with p_H . This cannot go on indefinitely. y must ultimately reach another

critical value; at this value of y , π must be equal to $(1 - \pi N_H)/N_L$; both p_H and p_L must be equal to y ; and any further reduction in y will make p_L greater than y . A simple calculation establishes that the critical value of y must be $L/(N_H + N_L)$ and the corresponding value of π must be $1/(N_H + N_L)$.

Once y falls below $L/(N_H + N_L)$, the constraint $p_L \leq y$ will bind, and therefore there is nothing to be gained by further lowering π . It is easily checked that it is optimal to set $p_L = p_H = y$ and to raise π to meet the IRL constraint (since $\pi > 1/(N_H + N_L)$) and $p_L = p_H$, ICH cannot bind).

The value of π as a function of y in this region of the parameter space will be (from IRL) $\pi = (L - N_L y)/N_H L$. Now as y goes to 0, this value of π goes to a number greater than 1. Therefore, y must hit a critical value beyond which reducing y does not increase π . This value of y is $L(1 - N_H)/N_L$. Below this value of y , $\pi = 1$.

Compiling all the results proved above, we have the claimed result. Proved

We next observe that at the solution to [mb] the suppressed constraint ICL does not always bind.

CLAIM A2. ICL binds at the values of p_H, p_L, π, T_H , and T_L which solve the [mb] iff (a) if $F \geq L$, and $y < L$, and (b) if $F < L$ and $y < L \cdot [N_H + N_L]^{-1}$.

Proof. Immediate from substitution of the solution of [mb] into ICL.

The next step is to note that since when ICL does not bind [MB] is the same as [mb], the solution to [MB] is just the solution to [mb] when conditions (a) and (b) do not hold. We state this as

CLAIM A3. If $F \geq L$, and $y \geq L$, or if $F < L$ and $y \geq L \cdot [N_H + N_L]^{-1}$, the solution to [MB] is the same as the solution to [mb].

Finally, we directly solve the problem for the case where it is known that ICL binds, assuming that v/δ is not too large. The solution is given below (we only describe the solution for values of y higher than $L(1 - N_H)/N_L$ to prevent the statement from becoming too long—the full statement is given in the previous version of the paper).

CLAIM A4. Let $N_L/N_H > v/\delta$ and $v/\delta + N_H v/N_L \delta < 1$. Then the solution to [MB] for the parameter values $L(1 - N_H)/N_L \leq y < L$

if $F \geq L$ and $L(1 - N_H)/N_L \leq y \leq L \cdot [N_H + N_L]^{-1}$ if $F < L$, is as follows.

If $L > y \geq L \cdot [N_H + N_L]^{-1}$ and $L(1 - \nu/\delta) \leq F$, the outcome is $\pi = 1$, and T_H set to solve the equation $L - y - \delta T_H = 0$.

If $L > y \geq L \cdot [N_H + N_L]^{-1}$, $L \leq F < L(1 + \nu/\delta)$, the outcome is $\pi = y/L$ and $T_H = 0$.

If $L \cdot (1 - N_H)/N_L \leq y < L \cdot [N_H + N_L]^{-1}$, $L(1 + \nu/\delta) \leq F$, the outcome is $\pi = 1$ and T_H set to solve $L - y - \delta T_H = 0$.

If $L \cdot (1 - N_H)/N_L \leq y < L \cdot [N_H + N_L]^{-1}$, $L(1 + \nu/\delta) > F \geq L(\nu/\delta + N_H \nu/N_L \delta)$, the outcome is π and T_H set to solve $\pi L - y - \delta T_H = 0$ and $L(1 - N_H \pi)/N_L = y$.

If $L \cdot (1 - N_H)/N_L \leq y < L \cdot [N_H + N_L]^{-1}$, $F < L(\nu/\delta + N_H \nu/N_L \delta)$, the outcome is $\pi = (N_H + N_L)^{-1}$ and $T_H = 0$.

Proof. Note that since ICH does not bind, raising p_H is always a good thing. Therefore, $p_H = y$. Assume now that $T_H > 0$, and consider the effect of a ΔT_H reduction in T_H on the bureaucrat's objective function. To keep ICL satisfied, we must reduce either p_L or π . In the case when we reduce p_L , the gain is $\nu N_H \Delta T_H$ which is less than the loss that is $N_L \delta \Delta T_H$ by our condition $\nu/\delta < N_L/N_H$. Therefore, it will never pay to reduce p_L . In fact, p_L will be raised until either IRL binds or $p_L = y$.

Assume next that IRL binds. This combined with ICL implies that

$$(A2) \quad \pi L - y - \delta T_H = 0.$$

From (A2) $d\pi/dT_H = \delta/L$. Using this in combination with the formula for $dp_L/d\pi$ derived from IRL, we find that an increase in T_H (weakly) increases the bureaucrat's welfare if $F \geq (1 + \nu/\delta)L$. Therefore, if $F \geq (1 + \nu/\delta)L$, an increase in π accompanied by the corresponding rise in T_H must increase the bureaucrat's welfare. Conversely, as long as $p_L < y$, if $F < (1 + \nu/\delta)L$, a reduction in T_H must raise the bureaucrat's welfare.

Next let IRL not bind. Then from ICH, $dT_H/d\pi = L(1 + N_H/N_L)/\delta$. Therefore, an increase in π accompanied by a rise in T_H (weakly) raises the bureaucrat's welfare iff $F \geq L(\nu/\delta + \nu N_H/N_L \delta)$.

Since $L(\nu/\delta + \nu N_H/N_L \delta) < L(\nu/\delta + 1)$, $F \geq L(\nu/\delta + 1)$ suffices in both cases. Therefore, under this condition π will be set equal to 1 (since an increase in π accompanied by an increase in T_H increases the bureaucrat's welfare). Therefore, $p_L = \min\{(1 - N_H)/N_L, y\}$ which, given our restriction on y , means that $p_L = (1 - N_H)/N_L$.

Next consider the case where $L(v/\delta + vN_H/N_L\delta) \leq F < L(v/\delta + 1)$. In this case it does not pay to increase π once IRL binds but so long as IRL does not bind, π will be increased. Therefore, either $\pi = 1$, or π must be such that IRL just binds. But if IRL does not bind, we must have $p_L = y$ which along with $\pi = 1$ implies that IRL is violated (as long as $y \geq (1 - N_H)/N_L$). Therefore, IRL must bind; i.e., we must have $L(1 - N_H\pi)/N_L = p_L$.

Now we know from above that when IRL binds and $p_L < y$, if $F < (1 + v/\delta)L$ the bureaucrat always wants to reduce T_H . Therefore, at the optimum we will have $T_H = 0$. This implies that the optimal values of π and p_L will be, respectively, y/L and $L(1 - N_H y/L)/N_L$.

By contrast, when $y < L/(N_H + N_L)$, solving IRL and ICL with $T_H = 0$ yields a solution for p_L which is greater than y . Therefore, we must choose $T_H > 0$. Specifically, we will choose $p_L = y$ and π and T_H to satisfy $\pi L - y - \delta T_H = 0$ and $L(1 - N_H\pi)/N_L = y$.

Proved

Claims A3 and A4 between them describe the full solution to the bureaucrat's problem [MB].

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

- Acemoglu, Daron, "Reward Structures and the Allocation of Talent," mimeo, 1992.
- Aghion, Philippe, and Robin Burgess, "Financing Development in Eastern Europe and the Former Soviet Union," in A. Giovannini, ed., *Finance and Development: Issues and Experience* (Cambridge, UK: Cambridge University Press, 1993).
- Banerjee, Abhijit V., "Eliminating Corruption," mimeo, MIT, 1995.
- Becker, Gary, and George Stigler, "Law Enforcement, Malfeasance and the Compensation of Enforcers," *Journal of Legal Studies*, III (1974), 1-19.
- Bolton, Patrick, and Gérard Roland, "Privatization Policies in Central and Eastern Europe," *Economic Policy*, XV (1992), 276-309.
- Cadot, Olivier, "Corruption as a Gamble," *Journal of Public Economics*, XXXIII (1987), 223-44.
- Che, Yeon-Koo, and Ian Gale, "Auctions with Financially Constrained Bidders," mimeo, University of Wisconsin at Madison, 1994.
- Clague, Christopher, "Rule Obedience, Organizational Loyalty, and Economic Development," *Journal of Institutional and Theoretical Economics*, CXLIX (1993), 393-414.
- Dixit, Avinash, *The Making of Economic Policy* (Cambridge, MA: MIT Press, 1996).
- Holmstrom, Bengt, and Paul Milgrom, "Multi-Task Principal Agent Analyses," *Journal of Law, Economics and Organizations*, VII (1991), 24-52.
- Kofman, Alfredo, and Jacques Lawarree, "On the Optimality of Allowing Corruption," mimeo, 1990.
- Laffont, Jean-Jacques, and Jean Tirole, *A Theory of Incentives in Procurement and Regulation* (Cambridge, MA: MIT Press, 1993).
- Leff, Nathaniel H., "Economic Development through Bureaucratic Corruption," in

- Monday Ekpo, ed., *Bureaucratic Corruption in Sub-Saharan Africa* (Washington, DC: University Press of America, 1979).
- Lewis, Tracy, and David Sappington, "Assigning Productive Assets and Profit-Sharing Opportunities to Wealth Constrained Agents," mimeo, University of Florida, 1996.
- Mauro, Paolo, "Corruption, Country Risk, and Growth," *Quarterly Journal of Economics*, CX (1995), 681–713.
- Murphy, Kevin M., Andrei Shleifer, and Robert Vishny, "Why Is Rent-Seeking so Costly to Growth?" *American Economic Review Papers and Proceedings*, LXXXIII (1993), 409–14.
- Nye, Joseph S., "Corruption and Political Development: A Cost-Benefit Analysis," in Monday Ekpo, ed., *Bureaucratic Corruption in Sub-Saharan Africa* (Washington, DC: University Press of America, 1979).
- Olson, Mancur, "Dictatorship, Democracy and Development," *American Political Science Review*, LXXXVIII (1993), 567–76.
- Sah, Raaj K., "Social Osmosis and Patterns of Crime," *Journal of Political Economy*, XCIX (1991), 1272–95.
- Shleifer, Andrei, and Robert Vishny, "Corruption," *Quarterly Journal of Economics*, CVIII (1993), 599–617.
- Tirole, Jean, "Collusion and the Theory of Organisations," in *Advances in Economic Theory: Proceedings of the 6th World Congress of the Econometric Society*, Jean-Jacques Laffont, ed., (Cambridge, UK: Cambridge University Press, 1992).
- . "A Theory of Collective Reputations," *Review of Economic Studies*, LXIII (1996), 1–22.
- Waterbury, John S., "Endemic and Planned Corruption in a Monarchical Regime," in Monday Ekpo, ed., *Bureaucratic Corruption in Sub-Saharan Africa* (Washington, DC: University Press of America, 1979).
- Weitzman, Martin, "Is the Price System or Rationing More Effective in Getting a Commodity to Those Who Need it Most?" *Bell Journal of Economics*, VIII (1977), 517–24.
- Wilson, James Q., *Bureaucracy: What Government Agencies Do and Why They Do It* (New York: Basic Books, 1989).