

Inflexible Rules in Incentive Problems

By TRACY R. LEWIS AND DAVID E. M. SAPPINGTON*

In practice, contracts involve "standard terms" or "rules," allowing for variations only under "exceptional" circumstances. We develop a simple model in which optimal contracts display this feature, even in the absence of transactions costs. Rules arise when an agent has "countervailing incentives" to misrepresent private information. These incentives are created by endowing the agent with a critical factor of production ex ante. Applications in regulatory, labor, and legal settings are developed.

Consider the following stylized setting. An absentee landowner hires a tenant farmer to produce crops on her land. The amount of crop produced depends entirely on how extensively the soil is cultivated. Mules provide the power for cultivation. Mules are also employed to transport cargo in the village and to pull barges up the nearby river. There is a fixed supply of mules in the village, some of which are owned by the absentee landowner.

Due to randomness in the derived demand for mules, the price at which the mules' services can ultimately be secured for cultivation is a variable that can neither be predicted perfectly nor observed by the absentee landowner. One option available to the landowner is to sell her mules at their fair market value before she leaves the village, and instruct the farmer to secure the appropriate number of mules after their (rental) price becomes known. Alternatively, the landowner could place her mules at the dis-

posal of the tenant farmer. Such a policy might seem ill-advised, because the absentee landowner cannot determine whether her mules are being employed efficiently. The farmer might, for example, rent out the mules to the local barge haulers rather than use them to cultivate the land. Nevertheless, it turns out that the absentee landowner always endows the farmer with complete control over some of her mules. Doing so helps limit the farmer's incentive to exaggerate production costs, in a manner to be explained shortly.

In fact, the use of "capital grants" of this type is optimal much more generally. A "principal" (here, the landowner) will generally gain from endowing an "agent" (here, the farmer) with a productive resource (for example, mules) at the start of their relationship. Doing so helps limit the rents the agent can command from his private information about production costs. Without a capital endowment, the binding incentive problem for the principal is to prevent the agent from exaggerating realized production costs in an attempt to secure more generous compensation for his output. By endowing the agent with productive capacity up front, on the other hand, a countervailing incentive is created. Now, the agent has reason to understate the value of the resource placed at his disposal, thereby claiming that more generous compensation is required from other sources. Consequently, if the direct cost of working for the principal without benefit of the capacity (for example, the rental price of mules) is positively correlated with the

*Department of Economics, University of California-Davis, CA 95616, and Bell Communications Research, 435 South St., Morristown, NJ 07960-1961. We wish to thank two anonymous referees as well as Vincent Crawford, Joel Demski, Joseph Farrell, Drew Fudenberg, Theodore Groves, Robert Inman, Preston McAfee, Meg Meyer, Roger Myerson, John Panzar, Martin Perry, Andrew Postlewaite, Michael Riordan, William Rogerson, David Sibley, Richard Schmalensee, Suzanne Scotchmer, Al Slivinski, Joseph Stiglitz, Jean Tirole, and seminar participants at Bellcore, Berkeley, Duke, Florida, MIT, Northwestern, Penn, Vanderbilt, and Western Ontario for very helpful comments in preparing this paper. The views expressed here are not necessarily those of Bellcore.

(shadow) value of the capacity (as in the landowner-farmer example, where there is perfect correlation), the agent's normal incentive to exaggerate production costs will be offset to some extent by the countervailing incentive to understate these costs. As a result, the principal's task of motivating truthful disclosure of private cost information is eased when she endows the agent with productive capacity.

When countervailing incentives are created via capital endowments, the optimal contract between principal and agent exhibits some interesting features. Most importantly, the optimal contract is characterized by "rules" rather than "discretion" for a range of intermediate cost realizations. In other words, the agent is induced to produce the same level of output for many different cost levels. This insensitivity of performance to realized cost stands in stark contrast to the fine tailoring of performance to costs that is observed in optimal contracts when the agent's incentive is unilaterally to exaggerate production costs.¹

Rules arise as a natural resolution of the conflict between the agent's induced incentives to exaggerate low cost realizations and to understate high cost realizations. They do not arise from such considerations as bounded rationality and transactions costs (Oliver Williamson, 1975), nor are they due to risk aversion in our model.² Thus, the optimality of capital endowments stands as

¹Formally, the equilibrium is a fully separating one in most agency models, whenever a common regularity condition is imposed (see Section I). This observation is made in a general theoretical model by Roger Guesnerie and Jean-Jacques Laffont (1984), and in a variety of other settings, including the regulation of public utilities (David Baron and Roger Myerson, 1982), the provision of public goods (Theodore Groves and John Ledyard, 1985), optimal taxation (James Mirrlees, 1971), the resolution of labor disputes (Vincent Crawford, 1985), and the specification of debt contracts in credit markets (Joseph Stiglitz and Andrew Weiss, 1981).

²Furthermore, rules are not imposed (as in Jean Tirole, 1986) to limit the gains from collusion that might otherwise accrue to subordinates in a hierarchical relationship, or because of the inability of the principal to commit to a multiperiod incentive scheme (as in Baron and David Besanko, 1987, or Laffont and Tirole, 1986).

an independent explanation for why contracts may specify rules rather than afford considerable discretion to a privately informed agent.

Our analysis predicts that rules rather than discretion will be optimal in a variety of settings. Consider the regulatory setting, for example. Here, the regulator might authorize the purchase of specialized equipment or resources that will enable the regulated firm to participate in unregulated markets. Such participation can create countervailing incentives for the firm, which results in the same price being set in the regulated market for an entire range of cost realizations.

The "rule" in our model can also be interpreted as the absence of renegotiation of a legal contract that is signed by a buyer and supplier before supply conditions are known with certainty. Even though renegotiation of the initial agreement is feasible after all relevant supply information becomes known to the supplier, no renegotiation will occur under a broad range of circumstances.

A third application of our model is in the labor context. Most agency models of the labor setting presume there is an employer who designs a contract that a potential employee can either accept or reject. The employee has (or will acquire) private knowledge of his ability, and will accept the proposed terms of employment if and only if his net income on the job exceeds his reservation wage. Most importantly, this reservation wage is assumed to be independent of ability. In other words, the agent's expected earnings in other occupations is assumed to be unrelated to his talents in the principal's employ.³ While this may be a reasonable approximation in some employment situations, it will not be true in general.

Our model can be interpreted as one in which an employee learns his ability on the job after an initial training period. More able employees are generally able to command greater compensation in alternative

³Two other (rather different) models that allow the "agent's" "opportunity wage" to vary with his "ability" are those of Robert Feenstra and Tracy Lewis (1987) and John Moore (1985).

occupations; but the original employer has some control over the mobility of her employees. For example, a physician who takes on a new partner can limit the geographic region in which this latter individual can establish his own practice should he choose to withdraw from the partnership. Our analysis shows that although the employer will optimally limit the employee's mobility to some extent, she will never do so entirely. By structuring the employment contract to allow some mobility, the employer establishes a strictly positive correlation between a worker's ability and his outside opportunities. The countervailing incentives that result limit the rents the employee could otherwise command from his private information, and lead to wages that are insensitive to ability over some range of ability levels.

These three applications of our model are examined in detail in Section III. First, though, the general model is developed in Section I and our findings are stated formally in Section II. An intuitive explanation of our findings is also offered in Section II. Conclusions are drawn in Section IV, along with suggestions for further research. For expositional ease, the proofs of all formal results and the corresponding technical arguments are relegated to the Appendix.

I. Description of the Model

At the start of the game under consideration, the principal offers a contract to the agent. The agent can either accept the contract as it stands or reject it altogether. The interaction between principal and agent is not repeated.⁴ At the time she designs the contract, neither the principal nor the agent know the agent's marginal cost of production c . (See Figure 1.) Instead, the two parties know only that the agent's cost is drawn

⁴Thus, we follow most of the literature by endowing the principal with Stackelberg leadership abilities. This setup is plausible, for example, when the principal is a monopsonist and the agent is one of many potential workers with identical observable characteristics. Although the formal analysis proceeds in a nonrepeated setting, dynamic interpretations of the model are readily constructed.

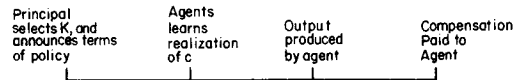


FIGURE 1. TIMING IN THE MODEL

from a population with frequency $f(c)$. $f(\cdot)$ is continuous and strictly positive on the closed interval $[c, \bar{c}]$. The following regularity conditions are also assumed to hold:

$$(RC) \quad \frac{d}{dc} \left\{ c + \frac{F(c)}{f(c)} \right\} \geq 0 \quad \text{and}$$

$$\frac{d}{dc} \left\{ c - \frac{[1 - F(c)]}{f(c)} \right\} \geq 0 \quad \forall c \in [c, \bar{c}].$$

These conditions are commonly imposed in the agency literature. They generally ensure the agent's observable performance varies strictly with his unobservable cost under the optimal contract. As we demonstrate in Section II, they do not ensure this property in our model.⁵

The key distinction between our model and others in the literature is the fact that we allow the principal to purchase a productive resource at the outset of her relationship with the agent. We call this resource "capacity," K . This capacity is placed at the disposal of the agent, provided the agent remains in the employ of the principal (as was the case with the mules in our introductory example).

To keep the analysis manageable, we presume a very simple fixed-proportions production technology: each unit of capacity enables the agent to produce exactly one unit of output for the principal at no cost. Thus, if the agent produces fewer units of output, Q , than the number of units of capacity he is given, K , the production will

⁵These conditions are imposed to dramatize our conclusions about "pooling" (i.e., rules rather than discretion): even when the conditions that generally ensure a fully separating equilibrium (i.e., full discretion) are imposed, pooling (i.e., inflexible rules) is optimal in our model. Notice that both regularity conditions are satisfied by the uniform distribution, for example.

cost him nothing.⁶ Furthermore, the agent can employ the excess capacity, $K - Q$, as he sees fit, since the principal cannot monitor his use of the capacity. In particular, the agent can use the capacity to produce in different "markets," where the capacity is a substitute for other factors of production. (Recall the farmer in our introductory example could employ or rent the mules to pull barges or transport cargo.)

The more costly are these substitute factors of production, the greater the (shadow) value of capacity in alternative uses. The prices of these substitute inputs are assumed to be positively correlated with c , the unit cost of producing output ($Q > K$) for the principal. In other words, input prices are correlated across "markets." For simplicity, we assume the unit (shadow) value of capacity is perfectly correlated with the unit cost of producing output for the principal; that is, c represents both these levels.

The principal's objective is to maximize her net expected gain from dealing with the agent. She values output, Q , according to the increasing concave function, $V(Q)$. Payments from the principal to the agent are denoted $P(\cdot)$, and the *ex ante* unit cost of capital to the principal is c_e . Any capacity the principal wishes to install must be in place well in advance of production. In particular, K must be chosen before the agent learns the realization of c .⁷ For the purpose of normalization, we assume c_e is the expected unit value of capacity in subsequent use. Thus, if the principal decides to install any capacity, she does so purely for incentive purposes; she does not anticipate any

⁶Of course, the unit cost of producing output in excess of K units is c .

⁷An alternative setting would have capacity determined after the agent learns the realization of c . Although preliminary investigation suggests that countervailing incentives and "rules rather than discretion" will continue to characterize the optimal incentive contract in this alternative setting, we focus on the case we do because the timing is more natural in a variety of settings. For example, in the labor interpretation of our model, the employee's mobility may be determined by the specificity of his training; and the training course must be completed before the employee learns his ability on the job.

direct profit from her investment. The principal's utility is given by $U(Q, P, K) = V(Q) - P - c_e K$.

The agent's goal is to maximize his profit, π . His compensation could conceivably come from two sources: from payments made by the principal, and from gains derived from the use of the installed capacity. Thus, the agent's profit when he is paid P to produce Q , when K is the level of installed capacity, and when marginal production cost (and unit value of capacity) c is realized is $P - cQ + cK$.

To convince the agent to work for her, the principal must promise the agent at least his reservation profit level, which is normalized at zero. Furthermore, we assume that after he learns his marginal cost of production and the unit value of capacity, the agent always has the option of terminating his relationship with the principal without penalty.⁸ Upon doing so, the agent will pursue his most profitable alternative opportunity, which has an expected value of zero.

To review the timing in the model (see Figure 1), the principal and agent initially share the same beliefs about c . At this time, the principal chooses capacity, K , and announces the terms of compensation. Later, the agent learns his unit cost of production and the unit value of capacity. Then the agent chooses how much output to produce and how to employ the capacity. Finally, payments are made as per the terms of the agreement.

To state the contracting problem formally, we appeal to the revelation principle (for example, Roger Myerson, 1979), which ensures that without loss of generality, the contract can be implemented by conditioning the output and associated transfer payment on the agent's truthful report of c . Thus, we can think of the principal as speci-

⁸For simplicity, we assume *no* penalty can be imposed on the agent if he withdraws from the contract. (Alternatively, the wealth-constrained agent can post *no* bond with the principal.) The qualitative conclusions we report below continue to hold provided feasible penalties (or bonds) are sufficiently small relative to the maximum expected total surplus that could arise from the relationship between the principal and agent.

fyng a menu of possible output levels, $Q(c)$, and associated payments, $P(c)$, indexed by the agent's truthful announcement of costs. The contracting problem [CP] is

$$\text{Maximize}_{P(\cdot), Q(\cdot), K} \int_{\underline{c}}^{\bar{c}} [V(Q(c)) - P(c)] \\ \times f(c) dc - c_e K$$

subject to $\forall c, \hat{c} \in [\underline{c}, \bar{c}]$:

$$\text{(IR)} \quad \pi(c|c) \geq 0, \text{ and}$$

$$\text{(IC)} \quad \pi(c|c) \geq \pi(\hat{c}|c),$$

where $\pi(\hat{c}|c) = P(\hat{c}) - cQ(\hat{c}) + cK$.

The individual rationality (IR) constraints ensure the agent is always guaranteed at least his reservation level of profit. For simplicity, we assume $Q(c) > 0 \forall c \in [\underline{c}, \bar{c}]$ in the solution to [CP]. Thus the principal designs the contract so as to always retain the services of the agent, and have him produce some output. Notice that the (IR) constraints also ensure the agent expects to receive at least his reservation profit of zero when he signs the contract initially. The incentive compatibility (IC) constraints identify $Q(c)$ as the output the agent will produce and $P(c)$ as the corresponding payment he will receive when c is the realized level of costs. Put differently, when c is the realized cost level, the agent will (weakly) prefer to truthfully reveal this fact rather than claim that some other cost level, \hat{c} , was realized.

If the principal could observe the realization of c along with the agent, she would implement the "first-best" policy. In this policy, the firm is instructed to produce the level of output, $Q^*(c)$, at which the marginal benefits and marginal costs of production are equated, that is, $V'(Q^*(c)) = c$. Furthermore, the agent would be paid the minimum required to induce this level of output; that is, $P^*(c) = c[Q^*(c) - K]$. Since there are no incentive problems in this hypothetical setting, the principal would have no strict preference to install any capacity up front.

In the setting under consideration, however, the realization of c is observed only by the agent, so the first-best policy will not be feasible. The optimal policy is described in Section II.

II. Properties of the Solution to [CP]

As a preliminary step in solving [CP], we record two properties that must hold in any incentive-compatible contract, that is, one that satisfies the (IC) constraints in [CP].

LEMMA 1: *Suppose $\pi(c) \equiv \pi(c|c)$ and $Q(c)$ are differentiable almost everywhere in the solution to [CP]. Then necessary and sufficient conditions for (IC) to hold are (i) $\pi'(c) = K - Q(c)$, and (ii) $Q'(c) \leq 0$, for almost all $c \in [\underline{c}, \bar{c}]$.*

Condition (ii) turns out to have particularly important implications for the solution to [CP], which is described in Proposition 1. The condition states that, in equilibrium, the agent will be induced to produce more output the lower are realized costs. If an attempt were made to induce greater output from an agent with higher costs, the terms of compensation would have to be sufficiently generous that an agent with slightly lower costs would always prefer to produce the larger level of output also.

PROPOSITION 1: *Suppose $\pi(c)$ and $Q(c)$ are differentiable almost everywhere in the solution to [CP]. Then the solution exhibits: (i) Rules rather than Discretion: that is, $Q(c) = Q^k > 0 \forall c \in [c_1, c_2]$, where $\underline{c} < c_1 < c_2 < \bar{c}$; (ii) Installation of Capacity: that is, $K = Q^k = Q^*(c_k)$, where $c_k \in [c_1, c_2]$; (iii) Rent Limitation: that is, $\pi(c) = 0 \forall c \in [c_1, c_2]$, although $\pi(c) > 0$ for $c \in [\underline{c}, c_1]$ and for $c \in (c_2, \bar{c}]$; (iv) Underproduction: that is, $Q(c) < Q^*(c) \forall c \in (\underline{c}, c_k)$; (v) Overproduction: that is, $Q(c) > Q^*(c) \forall c \in (c_k, \bar{c})$; and (vi) Efficiency at the Extremes: that is, $Q(\underline{c}) = Q^*(\underline{c})$, and $Q(\bar{c}) = Q^*(\bar{c})$.*

The solution to [CP] is illustrated in Figures 2 and 3. Properties (i) and (ii) of Proposition 1 reveal the two features of the solution that warrant particular emphasis.

First, there is always a nondegenerate intermediate range of cost realizations, $[c_1, c_2]$, in which rules rather than discretion prevail: the agent is induced to produce the same level of output for all cost realizations in this range. Second, the principal always installs some strictly positive level of capacity at the start of her relationship with the agent.

To understand these and the other features of the solution to [CP], first consider the incentive problem the principal would face if she did not install any capacity at all. In this case, the agent would always be tempted to exaggerate costs in an attempt to convince the principal that more generous compensation for production is in order. When the agent is afforded control over some positive level of the productive resource, however, a countervailing incentive is introduced. Now if the agent exaggerates the realization of c , he effectively exaggerates the value of the installed capacity, and thereby overstates his personal gain from the resource. The agent would like to claim that production costs are high and the value of the capacity is small; but he cannot do both simultaneously because the two measures are positively correlated. Thus, by installing capacity (which is a substitute in production for other costly inputs), the principal creates an incentive for the agent to understate c . This countervailing incentive helps mitigate the agent's normal incentive to exaggerate c , and thereby reduces the rents the agent commands from his private information.^{9,10}

⁹Notice that if production costs and capacity value were *negatively* correlated, the principal would not choose to install any capacity, K . With negative correlation, the normal incentive to exaggerate production costs would be exacerbated if K were strictly positive because a report of high costs would also imply that the value of the installed capacity is small.

¹⁰Ours is not the only model in the literature where countervailing incentives arise. In Peter Cramton and Thomas Palfrey (1986), for example, firms in a cartel may have an incentive to exaggerate costs in order to secure greater compensation if called upon to produce. They may also have an incentive to understate costs so as to exaggerate profits they forego when called upon to limit production. Similarly, in Cramton, Robert Gibbons, and Paul Klemperer (1987), joint owners of an asset will be tempted to exaggerate their true valuation

Notice from property (ii) of Proposition 1 that the amount of capacity installed will be limited. If K is chosen to be too large, the agent's incentive will always be to understate c in an attempt to convince the principal that the right to control K is not a very lucrative privilege. The key point is that the principal gains from the simultaneous presence of countervailing incentives, and she will install a moderate level of capacity to ensure this presence. In particular, K will be chosen so that the binding incentive problems are to prevent exaggeration of low realizations of c and to limit understatement of high realizations of c .

The incentive to exaggerate low cost realizations is optimally mitigated by reducing the output the agent is called upon to produce when he reports a high realization of c . (See Figure 2.) The smaller the level of output, the smaller the difference in total costs of production associated with two distinct levels of marginal cost. Consequently, the potential gain in profit from exaggerating c is reduced when induced production levels, $Q(c)$, are set below the corresponding efficient levels, $Q^*(c)$. As indicated in property (iv) of Proposition 1, the output induced for all realizations of $c \in (c_l, c_k)$ will fall short of efficient levels.¹¹ This is the region of low

of the asset if they are to sell their share, but will want to understate their valuation if they are to purchase other shares. Also, in Michael Riordan and Sappington (1987), a principal may prefer to link tasks for an agent when the costs of performing the tasks are negatively correlated. With negative correlation, the normal incentive to exaggerate costs of performing the first task is somewhat mitigated because such exaggeration amounts to a prediction that costs will be particularly low in subsequent tasks.

¹¹The optimal level of output, $Q(c)$, for $c \in [c_l, c_1]$ is given by $V'(Q(c)) = c + [F(c)/f(c)]$. Notice that the efficient level of output will be induced when the smallest possible cost level, c_l , is realized. This is because deviations from efficient performance, $Q^*(c_l)$, would be instituted only to reduce the agent's gains from a false report of c_l . But there are no cost realizations below c_l that the agent could exaggerate by reporting c_l . Hence, no output distortion will be implemented for the smallest cost realization. Analogous logic explains why the efficient level of output will also be induced for the highest cost realization, \bar{c} , as reported in property (vi) of Proposition 1.

cost realizations in which the dominant incentive is for the firm to exaggerate c .

In the complementary region of high realizations of $c \in (c_k, \bar{c})$, the dominant incentive for the agent is to understate the true value of c . To prevent such understatement in this region, output is optimally expanded beyond efficient levels, as reported in property (v) of Proposition 1 and illustrated in Figure 2. The additional output is more costly to produce when costs are truly high; hence, the potential profit gain from understating c is reduced.¹²

The cost realization labeled c_k defines the "boundary" between the two regions of countervailing incentives. To best control the agent's incentive to exaggerate cost realizations below c_k , the principal would like to continue to induce output below the efficient level for all $c < c_k$, along the extension of segment AB in Figure 2. And to limit the agent's tendency to understate the higher realizations of c , the principal would like to continue to induce output above the efficient level for all $c > c_k$, along the extension of segment $A'B'$ in Figure 2. To do so, however, would require that for some range of cost realizations around c_k , larger output levels would be associated with higher cost realizations (that is, the solution to [CP] would have $Q'(c) > 0$ in some range). But we know from Lemma 1 that such an arrangement is not feasible.

Consequently, the principal can do no better than to institute rules rather than discretion in the region $[c_1, c_2]$ that surrounds c_k . In other words, the agent is instructed to produce the same level of output, Q^k , for all cost realizations in this region, as stated in property (i) of Proposition 1 and illustrated in Figure 2.

Of course, the reason the principal institutes any output distortions and installs any capacity is to limit the rents the agent commands from his private information. The pattern of rents that does accrue to the agent in the solution to [CP] is characterized in condition (i) in Lemma 1 and in property

¹²The optimal level of output, $Q(c)$, for $c \in [c_2, \bar{c}]$ is given by $V'(Q(c)) = c - [1 - F(c)]/f(c)$.

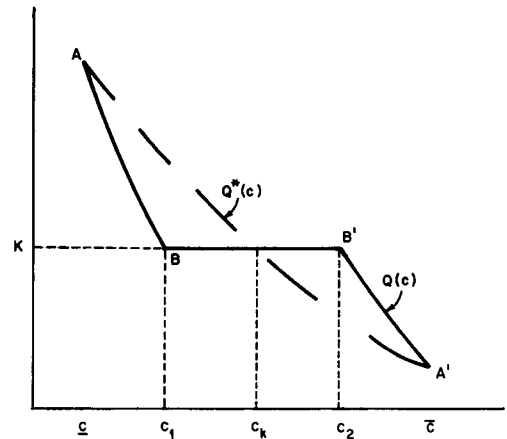


FIGURE 2. OPTIMAL VS. EFFICIENT OUTPUT LEVELS

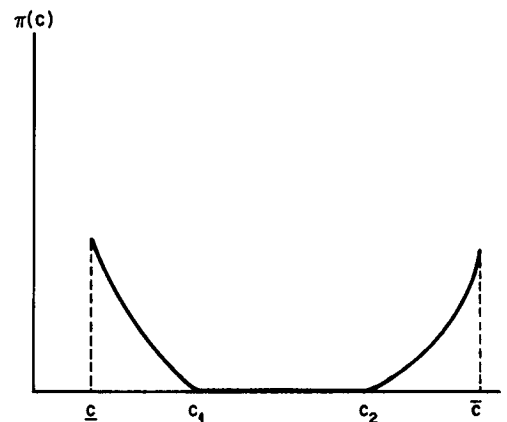


FIGURE 3. REALIZED PROFIT IN THE SOLUTION TO [CP]

(iii) of Proposition 1; the pattern is also illustrated in Figure 3. In the region of rules rather than discretion, $[c_1, c_2]$, the agent earns no rents. With installed capacity equal to Q^k , variations in c do not affect the agent's profit in this region. A higher c implies higher production costs; but it also implies a greater value of the installed capacity, and the two effects are exactly offsetting.¹³

¹³The linearity in our model is what ensures that profits do not vary with c in the region $[c_1, c_2]$. If the unit value of capacity were an increasing, strictly concave function of c , then the (IR) constraint would bind

In the regions where discretion is afforded the agent, he will receive some rent from his private information. The agent can secure profits when small cost levels $c \in [\underline{c}, c_1)$ are realized because of the inherent cost advantage he enjoys in this case. Profits also arise for the larger realizations of $c \in (c_2, \bar{c}]$ because the value of the installed capacity is particularly great in those events. Of course, the principal could reduce the rents associated with high realizations of c if she did not install any capacity up front. However, providing the agent with capacity introduces countervailing incentives, which limit the rents associated with smaller cost realizations. On balance, the principal always prefers to install a strictly positive level of capacity.¹⁴ The optimal choice of $K = Q^k$ is given by equation (A6) in the Appendix. This choice takes into account the effect of Q^k both on net expected surplus in the region of rules $[c_1, c_2]$ and on the agent's expected rents in regions $[\underline{c}, c_1]$ and $[c_2, \bar{c}]$.

Before proceeding to discuss some applications of our theoretical model in Section III, we briefly present an example to demonstrate that the qualitative effects we have identified can be of significant magnitude. In Figure 4, we illustrate the solution to [CP] for the case where marginal costs always lie between 1 and 2, and all cost realizations in this interval are initially thought to be equally likely, that is, $f(c) = 1 \forall c \in [1, 2]$. It is also

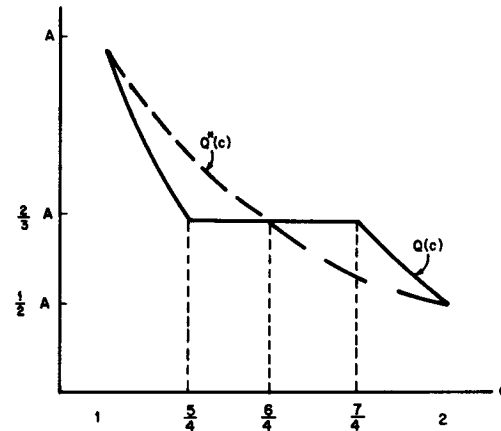


FIGURE 4. SOLUTION WITH UNIFORM BELIEFS AND $V(Q) = A \ln Q$

assumed that $V(Q) = A \ln Q$, where $A > 0$ is a constant. In this case, efficient output levels range between $\frac{1}{2}A$ and A (that is, $Q^*(1) = A$ and $Q^*(2) = \frac{1}{2}A$). Optimal output levels are given by

$$Q(c) = \frac{A}{2c-1} \quad \forall c \in [1, \frac{5}{4}] \quad \text{and by}$$

$$Q(c) = \frac{A}{2[c-1]} \quad \forall c \in [\frac{7}{4}, 2].$$

For the intermediate cost realizations $[5/4, 7/4]$, induced output, Q^k , will be $\frac{2}{3}A$, which is also the level of installed capacity. Thus, for 50 percent of the possible cost realizations, the same output level is induced. Somewhat loosely, then, rules will prevail over discretion with probability one-half in this example.¹⁵

in the solution to [CP] for exactly one realization of c . However, even in this case, there would be a nondegenerate region in which the same level of output would be induced for all cost realizations. Thus, even when neoclassical production technologies and/or differential returns to capacity in different uses are admitted, rules rather than discretion will continue to characterize the optimal incentive contract. (For details of the nonlinear analysis, but one where K is not a choice variable of the principal, see Lewis and Sappington, 1988.) We also suspect that the conclusion will continue to hold when unit production costs are positively but imperfectly correlated with the unit value of capacity.

¹⁴As it does here, the individual rationality constraint also binds for an intermediate rather than an extreme realization of the private information in the aforementioned multi-agent models of Cramton, Gibbons, and Klemperer (1987) and Cramton and Palfrey (1986).

¹⁵In fact, this conclusion is true more generally. It is straightforward to demonstrate that when $f(c)$ is a uniform density on any closed interval $[\underline{c}, \bar{c}]$, the length of the region in which rules prevail, $[c_1, c_2]$, will be exactly $\frac{1}{2}[\bar{c} - \underline{c}]$, regardless of the principal's preferences, $V(Q)$. This result follows because when the unknown parameter is uniformly distributed, the optimal quantity distortions to limit understatement and overstatement of c are symmetric. In particular, it follows

III. Applications of the Analysis

In this section, we examine the implications of our theoretical analysis in three institutional settings: regulating the activities

directly from the formulas given in fns. 11 and 12 that

$$\frac{dQ(\cdot)}{dc} = \frac{2}{V''(Q(\cdot))} \quad \text{for } c \in (\underline{c}, c_1) \quad \text{and}$$

$$\quad \quad \quad \text{for } c \in (c_2, \bar{c})$$

in the solution to [CP].

Thus, with identical slopes for the two declining segments of the optimal quantity schedule (see Figures 2 and 4), each being twice the slope of the efficient quantity schedule, $Q^*(c)$, the horizontal distance between the two segments is everywhere $\frac{1}{2}[\bar{c} - \underline{c}]$. Thus, even as the *ex ante* unit cost of capacity, c_e , rises above its expected value so that less capacity ($K > 0$) is purchased initially, induced output will be at this capacity level for 50 percent of the possible cost realizations.

With other distributions, changes in c_e may influence the optimal range of "rules." The direction of influence depends upon the relative slopes of the optimal quantity schedules for $c \in (\underline{c}, c_1)$ and for $c \in (c_2, \bar{c})$. When c_e rises, the principal purchases less capacity up front, and so "rules" prevail at a smaller level of output, \hat{Q}^k . If $|Q'(c)|$ is everywhere greater for $c \in (\underline{c}, c_1)$ than for $c \in (c_2, \bar{c})$ in the solution to [CP], then the horizontal distance between the two declining segments of the optimal quantity schedule is greater the smaller is Q . Thus, an increase in c_e will result in a larger region where rules prevail.

The relative slopes of the two declining segments of the optimal quantity schedule depend upon the distribution of c . Intuitively, as large realizations of c become more likely and small realizations of c become less likely, quantity distortions will decline for $c \in (\underline{c}, c_1)$ and will increase for $c \in (c_2, \bar{c})$. The explanation for these conclusions is the following. For $c \in (\underline{c}, c_1)$, the binding incentive problem is to prevent exaggeration of c . Output restrictions are instituted to limit such exaggeration, particularly for the higher realizations of c . But as these realizations become relatively more likely, distortions involve a greater expected cost, and so are reduced. On the other hand, for $c \in (c_2, \bar{c})$, the binding incentive problem is to limit understatement of c . Such understatement is best deterred by expanding output above its efficient level, particularly for the low c realizations. As these low realizations become relatively less likely, the expected costs of distortions decline, and so greater distortions are implemented.

These effects are readily demonstrated, for example, with a linear density function $f(c) = a + bc$ where $b \in [0, 1]$ and where $f(c) > 0$ for $c \in [0, 1]$. In this case, as b increases above zero, $|Q'(c)|$ decreases for $c \in (\underline{c}, c_1)$ and increases for $c \in (c_2, \bar{c})$ in the solution to [CP].

of a firm that operates in both regulated and unregulated markets; labor contracting involving mobility restrictions; and legal contracting with provisions for renegotiation. The regulatory application is the most straightforward, so we begin with that one.

A. Regulated Firms Operating in Unregulated Markets

Most regulated firms produce multiple products, some of which are sold in unregulated markets. For example, AT&T produces not only long-distance transmission service, but also such products as telephones and computers which are sold in unregulated markets. An important policy issue concerns the extent to which regulated firms should be allowed to operate in unregulated markets.

The model developed in Sections I and II can inform one dimension of this issue. To see this, call the "principal" the regulator here. The regulator's objective is to maximize the sum of: (i) tax revenue collected from the firm; and (ii) the expected consumers' surplus derived from the firm's (agent's) activities in the regulated market.¹⁶ The regulator can set prices and taxes in the regulated market. She can also control the regulated firm's participation in an unregulated market. She does so by authorizing the acquisition of specialized inputs (for example, technical equipment or expertise) that enable production of K units of the unregulated service. The unit profit from sales in the unregulated market is assumed to be directly proportional to realized marginal costs, c , in the regulated market. The critical aspect of this assumption is the positive correlation between production costs in the regulated market and the value of the installed capacity that enables participation in the unregulated market. Such correlation is likely when the same or similar inputs can be used to produce in the two markets.

The timing in this setting is exactly as in the general model. Initially, neither the regu-

¹⁶The conclusions cited below also hold if the regulator's objective is to maximize a weighted average of consumers' and producer's surplus, provided the weight on the former exceeds the weight on the latter.

lator nor the firm know the realization of c . At this time, the regulator makes a long-term capacity decision that dictates the extent of participation in the unregulated market. She also specifies a menu of two-part tariffs that may be established in the regulated industry. After (privately) observing the realization of c , the firm chooses from the menu the two-part tariff it most prefers, and serves all demand that is forthcoming.¹⁷ The firm also maximizes its profit by selling all it can produce in the unregulated market.

The conclusions of Section II explain why the regulator will always choose to allow participation in the unregulated market even though no independent gains are anticipated from the venture, that is, even though the cost of capacity is equal to its expected value in subsequent use. Participation in unregulated markets is valuable because it creates countervailing incentives. The firm's natural incentive to exaggerate production costs in the regulated market is partially offset by the incentive that is created to understate the value of participation in the unregulated market. Under the optimal regulatory policy, the dominating incentives are for the firm to exaggerate low-cost realizations and to understate high realizations of c . The former incentive is optimally mitigated by establishing usage prices in excess of marginal cost for low-cost realizations, $c \in (c_l, c_k)$ in Figure 5. The higher prices reduce output, and thereby reduce the potential profit gain from exaggerating costs. The latter incentive is best limited by inducing usage prices below marginal cost for the higher cost realizations, $c \in (c_k, \bar{c})$ in Figure 5. For intermediate cost realizations $c \in [c_1, c_2]$, the conflict between the countervailing incentives is optimally resolved by establishing a single price that is always charged in the regulated industry, regardless of the actual realization of c .

It is apparent from Figure 5 that prices in the regulated market are optimally designed to be somewhat insensitive to realized costs, as under "price-cap" as opposed to rate-of-

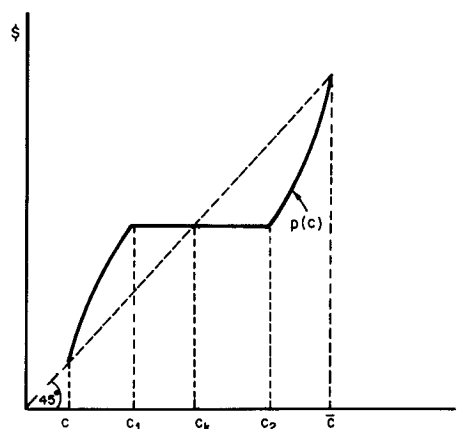


FIGURE 5. OPTIMAL REGULATED PRICE, $p(c)$

return regulation. What is perhaps of most interest about this price "stickiness" is the fact that it does not arise from risk-sharing concerns: it arises solely from considerations of incentives and rent distribution. And again, as indicated in fn. 13, the presence of a region of inflexible prices is *not* an artifact of the simple linear structure we have assumed for expositional clarity.

B. Employment Relations and Mobility Constraints

Our analysis is also relevant in the classic setting where an employer hires a worker from a pool of laborers with varying and unknown ability levels. The analysis is particularly applicable when the employer can influence the worker's ability to transfer the skills he learns on the job to alternative employment settings. Such influence may be effected through the extent of general versus idiosyncratic training provided, or through contractual covenants that explicitly restrict the worker's ability to pursue alternative employment options.

In this labor setting, c represents the agent's personal cost of performing the task in question. Thus, low values of c correspond to high ability levels on the job, and high values of c are associated with lower levels of ability. Initially, neither the employer nor the employee know the worker's

¹⁷We assume the regulator cannot observe the firm's operating costs, c , just as she cannot discern the shadow value of inputs used by the utility to produce in the unregulated market.

(that is, the employee's) ability on the job. At the time they share the same imperfect knowledge of c , the employer designs and presents to the employee a contract which specifies how much the agent will be paid, $P(\cdot)$, as a function of how much output, $Q(\cdot)$, he produces. Eventually, the worker alone learns his ability on the job. He might do so, for example, upon completion of a job training program. After learning c , the worker chooses how much output to produce. Finally, he is paid for his efforts according to the original contract. As in our general model, it is not possible for the employee to post a bond with the employer.

The critical feature of this labor setting is a feature that is generally missing in related studies in the literature. We allow the agent's opportunity wage, $W(\cdot)$, to depend upon his ability. In other words, we assume that agents who are particularly skilled in the principal's employ are also able to generate greater compensation elsewhere in the economy (in the sector of the self-employed, for example). And to prevent the agent from "quitting," the employer must guarantee the worker at least his opportunity wage.

The second important feature of this labor setting is the employer's power to affect the sensitivity of the worker's opportunity wage to his ability on the job. We assume the employer can choose a mobility parameter, m , that specifies the rate at which the agent's opportunity wage, $W(\cdot)$, increases with his privately observed ability; that is, $|W_c(c, m)| = m$, where the subscript denotes the obvious partial derivative. If the employer selects $m = 0$, then the agent's opportunities elsewhere are insensitive to his ability in working for the employer. But if the employer selects $m > 0$, then the more highly skilled is the worker, the more lucrative will his alternative employment opportunities be.

In practice, employers often have some control over the mobility (and hence the opportunity wage) of their employees. To illustrate, restrictive work covenants are common in many professions. Those covenants limit the geographic areas in which, for example, a doctor who leaves a partnership can establish his own office. Realistically, though, employers do not have unbr-

dled authority to limit the mobility of their employees. We model this limited authority as follows. We assume the employer can choose any finite value for $m \geq 0$, and thereby control how sensitive are an employee's opportunities to his realized ability.¹⁸ However, the employer cannot reduce the expected *ex post* opportunity wage of any employee below its *ex ante* level, \bar{W} . Formally, we require

$$\int_c^{\bar{c}} W(c, m) f(c) dc \geq \bar{W} \quad \forall m \geq 0.^{19}$$

In other words, the employer cannot be placed in a position where his alternative opportunities are reduced in expectation as a result of agreeing to work for the employer.²⁰ Graphically (see Figure 6), the employer has complete control over the slope of the employee's linear individual rationality constraint; but any variation in the slope must be accompanied by a compensatory change in the intercept that leaves the worker with at least his initial expected reservation wage of \bar{W} . To illustrate, the employee's opportunity wage may be very sensitive to his realized ability absent any action by the employer. If the employer chooses to reduce the employee's mobility, she can offset the expected deleterious effects of the restriction by promising a lump-sum severance pay to the agent, independent of his performance on the job.

The properties of the optimal contract in this labor setting are analogous to those described in Proposition 1. Most importantly, the employer will structure the employment

¹⁸For simplicity, we presume the principal has significant latitude in structuring the agent's mobility through the choice of m . In practice, there will be costs associated with implementing different levels of mobility, and there may be natural bounds on the maximum level of mobility. We abstract from these costs and bounds to better focus our analysis on the incentive effects that are of greatest interest here.

¹⁹We assume \bar{W} is sufficiently large that the employee never finds himself in a position of negative wealth *ex post*.

²⁰Notice that such a restriction can be readily enforced by a court that shares the initial symmetric beliefs, $f(\cdot)$, about the ability of a typical employee.

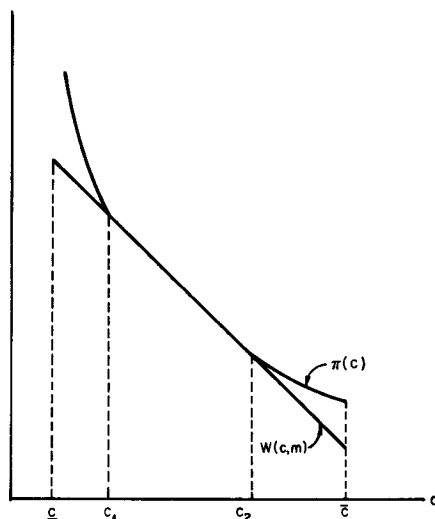


FIGURE 6. REALIZED PROFITS, $\pi(\cdot)$, AND OPPORTUNITY WAGE, $W(\cdot)$

contract so that more able workers have higher opportunity wages. In other words, the employer will choose *not* to limit the agent's mobility entirely, although she could do so by choosing $m = 0$. The reason the employer grants the worker some mobility by setting $m > 0$ is because the mobility introduces countervailing incentives. If the worker's opportunity wage were independent of his ability, the worker's incentive would always be to understate his ability in an attempt to convince the principal that output is more costly to produce than it really is. When his opportunity wage increases with his ability, however, the employee now has a reason to exaggerate his true ability. Since the worker with high ability can command a relatively high wage elsewhere, the employer will have to compensate the high-ability worker more generously to secure his services, *ceteris paribus*.

Thus, with $m > 0$, countervailing incentives are introduced. In equilibrium, the dominant incentives will be for the agent to exaggerate low realizations of ability and to understate high levels of ability. For the reasons described above, these incentives are optimally mitigated by inducing inefficiently large output levels when low levels of ability

are reported, and by implementing inefficiently small levels of output when high levels of ability are announced. The same output will be induced for an entire range of intermediate ability levels, $c \in [c_1, c_2]$, as an optimal response to the countervailing incentives that are created. In this range, the employee will receive exactly his opportunity wage, as illustrated in Figure 6. For the smaller and larger realizations of ability, however, the employee will receive some rents from his private information.

C. Renegotiation of Contracts

The final interpretation of the general model we offer pertains to the renegotiation of legal contracts. In this setting, our model suggests a rationale for why initial agreements are not always renegotiated even when significant changes occur in important environmental parameters.

The interaction considered in this interpretation is the following. Initially, a buyer finances the purchase of an inventory of Q^k units of a particular good. This inventory will be delivered to her by an identified supplier at a specified future date. The supplier's fee for acquiring and delivering the inventory is P^k . This initial agreement between buyer and supplier is legally binding upon both parties in the sense that either can costlessly enforce specific performance. Thus, unless the initial agreement is renegotiated, delivery of the Q^k units must be made in return for payment P^k .

Just prior to the delivery date, the supplier (alone) learns the unit value, c , of the inventory he has acquired. This unit value is the price at which the supplier could either sell or purchase additional units of the good in question if he were free to do so. After the supplier learns c , the initial agreement may be renegotiated. Because the value of the inventory is known only to the supplier, any renegotiation based on claims he makes must be incentive compatible. Furthermore, voluntary renegotiation implies that any change in the initial agreement must be beneficial to both parties. Formally this means that the new contract terms $\{Q(c), P(c)\}$ must guarantee both the supplier and buyer a higher

level of utility than would be secured under the original agreement, that is, $\pi(c) \geq P^k - cQ^k$, and $U(\cdot) \geq V(Q^k) - P^k$. (It is straightforward to verify that these inequalities are satisfied $\forall c \in [c, c_1]$ and $\forall c \in [c_2, \bar{c}]$ in the solution to [CP].)

Proposition 1 can be interpreted as describing the initial agreement and subsequent renegotiation schedule that are optimal from the buyer's point of view. The most important finding here is that the terms of the original agreement will be carried out for an entire range of realizations of $c \in [c_1, c_2]$. Only for extreme realizations of c will substantive renegotiation take place. When the inventory on hand turns out to be sufficiently valuable ($c > c_2$), the supplier is released from the original agreement to deliver Q^k . He is allowed to deliver a smaller quantity to the buyer and sell the remainder; but his compensation is also reduced so that the buyer is no worse off. When the supplier finds that additional inventory can be acquired sufficiently cheaply ($c < c_1$), increased delivery is agreed upon with a corresponding increase in compensation that leaves both the buyer and supplier better off.

The delivery level called for in the original agreement is *ex ante* efficient; that is, it is the level that maximizes the expected total surplus, given that $c \in [c_1, c_2]$. However, to limit the supplier's incentives to misrepresent his alternative opportunities, the renegotiated production level will generally not be efficient *ex post*. If c turns out to be unexpectedly high, the supplier will be tempted to downplay the profits he could earn by selling the inventory to others, so as not to reveal the true value of the asset in his possession. To counteract this tendency, delivery in excess of the efficient level is induced when $c \in [c_2, \bar{c}]$. On the other hand, the supplier's incentive will be to exaggerate unexpectedly small realizations of c , so as not to reveal how inexpensive it is to acquire additional units of the good. To mitigate this incentive, delivery below the efficient level is induced when the smaller realizations of c occur. For a range of intermediate c realizations, no renegotiation of the contract will occur, and the supplier will deliver Q^k in return for the agreed-upon payment, P^k .

Once again, then, it is optimal for the "principal" (buyer) to induce countervailing incentives for the "agent" (supplier). The buyer could specify zero output as the initial agreement (by setting $Q^k = 0$). If she did so, however, the supplier's incentive during renegotiation would always be to exaggerate the cost of purchasing the good; and motivating the supplier to perform when his incentive is unilaterally to exaggerate c is more difficult than when countervailing incentives exist. Once countervailing incentives are present, the optimal contract will entail a nontrivial region where rules rather than discretion prevail.

IV. Conclusions

We have developed a model in which optimal incentive contracts consist of standard terms or rules, with allowances for variations (i.e., discretion) only in exceptional cases. Contracts with this characteristic are easy to design and to administer, which may partially explain why they are observed in practice. We have shown, however, that even when transactions costs and bounded rationality are absent, optimal contracts may exhibit these same features. We have demonstrated how the presence of countervailing incentives will lead the principal to implement rules rather than discretion in the optimal contract in her effort to limit the agent's rents. Furthermore, we have explained why the principal will always choose to implement countervailing incentives.

The implications of our model were examined in three settings: (a) a regulated environment where the regulated firm may be authorized to participate in unregulated markets; (b) a labor setting where the employer can influence the mobility of her employees; and (c) a legal setting where renegotiation of an initial agreement is possible. In all three settings, countervailing incentives were induced and a region of inflexible rules characterized the optimal incentive contract.

The same will be true in other settings. For example, consider the case where the sponsor of prototype development must elicit information about the costs of producing the prototype that is developed. Our model sug-

gests that the sponsor can gain from endowing the developer (who is also the manufacturer) with some property rights to his prototype. Doing so mitigates the developer's incentive to exaggerate production costs, since high production costs imply that the prototype will be less valuable to other potential buyers, and so the developer's opportunity profit is somewhat meager. And with countervailing incentives in place, the optimal incentive contract will exhibit production plans that are relatively unresponsive to realized cost...and entirely unresponsive over some range.

In concluding, we suggest two avenues for future research. First, it would be helpful to identify additional factors that influence the optimal choice between rules and discretion. For example, it is not apparent how the availability of an imperfect monitor of the agent's costs would influence the principal's choice of an incentive scheme. How the principal would respond if her powers of intertemporal commitment were more limited would also be interesting to explore.²¹ Furthermore, the effect of competition among principals on optimal incentive contracts remains to be determined. Second, the examples developed in Section III suggest a rationale for sticky prices in the provision of certain goods and services. To our knowledge, the macroeconomic implications of incentive contracting in inhibiting price movements in the economy have received relatively little attention in the literature.

²¹When optimal contracts are characterized by rules rather than discretion, the losses that ensue in repeated settings from limited commitment ability on the part of the principal may be less pronounced. As Baron and David Besanko (1987) and Laffont and Tirole (1986) show, for example, if the principal cannot commit herself to ignore in future periods information that is gleaned from performance in the present period, she will be forced to induce performance that is not so finely tailored to the agent's realized ability. Losses in expected surplus result. But when the optimal contract with perfect intertemporal commitment ability already entails limited tailoring of performance to ability, the losses from limited commitment may be less pronounced.

APPENDIX

The proof of Lemma 1 follows directly from arguments presented in Baron and Myerson (1982) and Guesnerie and Laffont (1984), and so is omitted.

PROOF OF PROPOSITION 1:

First, it turns out to be useful to establish the following:

LEMMA 2: *The (IR) constraint binds over a single interval $[c_1, c_2]$, where*

$$\underline{c} \leq c_1 \leq c_2 \leq \bar{c}.$$

PROOF

Clearly the (IR) constraints in [CP] must bind for some c . Lemma 1 implies that $\pi''(c) \geq 0$ for almost all $c \in [\underline{c}, \bar{c}]$. Since the agent's profit schedule is weakly convex almost everywhere, (IR) binds at most along a single (possibly degenerate) interval of $[\underline{c}, \bar{c}]$.

For now we shall assume that the following condition, (M), is satisfied

$$(M) \quad K \in (Q^*(\bar{c}), Q^*(\underline{c})).$$

Later we will verify that (M) is in fact satisfied in the solution to [CP].

By Lemma 2, (IR) binds over some interval $[c_1, c_2]$. Therefore, it follows from part (i) of Lemma 1 that

$$(A1) \quad \pi(c) = \begin{cases} \int_c^{c_1} [Q(z) - K] dz & \text{for } c \in [\underline{c}, c_1]; \\ 0 & \text{for } c \in [c_1, c_2]; \\ \int_{c_2}^c [K - Q(z)] dz & \text{for } c \in [c_2, \bar{c}]. \end{cases}$$

It now proves useful to solve for $P(c)$. Since $\pi(c) = P(c) - c[Q(c) - K]$, we have

$$(A2) \quad P(c) = \begin{cases} \int_c^{c_1} [Q(z) - K] dz \\ \quad + c[Q(c) - K] & \text{for } c \in [\underline{c}, c_1]; \\ 0 & \text{for } c \in [c_1, c_2]; \\ \int_{c_2}^c [K - Q(z)] dz \\ \quad + c[Q(c) - K] & \text{for } c \in [c_2, \bar{c}]. \end{cases}$$

Substituting the expression in (A2) for $P(c)$ in the statement of [CP] and integrating by parts, we find after

some manipulation that the contracting problem [CP'] is to choose $\{K, Q(c), c_1, c_2\}$ in order to

$$\begin{aligned} \text{[CP']} \quad \text{Maximize} \quad & \int_{\underline{c}}^{c_1} \left\{ W(Q(c), c, K) \right. \\ & - \frac{F(c)}{f(c)} [Q(c) - K] \left. \right\} f(c) dc \\ & + \int_{c_1}^{c_2} W(K, c, K) f(c) dc \\ & + \int_{c_2}^{\bar{c}} \left\{ W(Q(c), c, K) \right. \\ & \left. + \frac{1 - F(c)}{f(c)} [Q(c) - K] \right\} f(c) dc, \end{aligned}$$

where $W(Q(c), c, K) \equiv V(Q(c)) + c[K - Q(c)] - c_c K$ is total surplus, $P(c)$ is given by (A2), and where (IC) requires (by part (ii) of Lemma 1)

$$\begin{aligned} \text{(A3)} \quad & Q(c) \geq K \text{ for } c \leq c_1; \text{ and} \\ & Q(c) \leq K \text{ for } c \geq c_2. \end{aligned}$$

Assuming an interior solution, pointwise maximization of [CP'] with respect to $Q(\cdot)$ yields the following:

$$\begin{aligned} \text{(A4a)} \quad & W_Q(\cdot) - \frac{F(c)}{f(c)} = 0 \\ & \text{for } c \in [\underline{c}, c_1]; \text{ and} \end{aligned}$$

$$\begin{aligned} \text{(A4b)} \quad & W_Q(\cdot) + \frac{1 - F(c)}{f(c)} = 0 \\ & \text{for } c \in (c_2, \bar{c}]. \end{aligned}$$

Let $H(c_1, c_2, K)$ be the maximal value of the objective function defined in [CP'] for given values of c_1, c_2 , and K . Applying Leibnitz's rule, we obtain

$$\begin{aligned} \text{(A5)} \quad \frac{\partial H(\cdot)}{\partial c_1} &= W(Q(c_1), c_1, K) \\ & - W(K, c_1, K) - \frac{F(c_1)}{f(c_1)} [Q(c_1) - K], \\ \frac{\partial H(\cdot)}{\partial c_2} &= W(K, c_2, K) - W(Q(c_2), c_2, K) \\ & - \frac{1 - F(c_2)}{f(c_2)} [Q(c_2) - K]. \end{aligned}$$

c_1 will be chosen so that $\partial H(c_1)/\partial c_1 \leq 0$. To characterize c_1 more precisely, define $\bar{c}_1 \equiv \inf\{c | Q(c) = K\}$, where $Q(c)$ satisfies (A4a). Clearly, $c_1 \leq \bar{c}_1$; for if c_1 were strictly greater than \bar{c}_1 , $Q(\cdot)$ would increase (discontinuously) with c at c_1 , which is ruled out by Lemma 1. Notice from (A5) that $\partial H(\bar{c}_1, \cdot)/\partial c_1 = 0$. For $c_1 < \bar{c}_1$, we have

$$\begin{aligned} \frac{d}{dc_1} \frac{\partial H(\cdot)}{\partial c_1} &= \left[W_Q(\cdot) - \frac{F(c)}{f(c)} \right] Q'(c) + W_c(\cdot) \\ &= K - Q(c) < 0, \end{aligned}$$

where the second inequality follows from (A4a). This implies that $\partial H(\cdot)/\partial c_1 > 0$ for $c_1 < \bar{c}_1$. Hence, $c_1 = \bar{c}_1$. A similar argument applied to c_2 establishes that $c_2 = \sup\{c | Q(c) = K\}$, where $Q(c)$ is characterized by (A4b).

To establish property (i) of Proposition 1, notice that property (M) implies

$$Q^*(\bar{c}) < K < Q^*(\underline{c}).$$

(We shall establish the optimality of (M) shortly.) Define c_k by $K = Q^*(c_k)$. Then, substituting for K in the expression above yields $Q^*(\bar{c}) < Q^*(c_k) < Q^*(\underline{c})$. Since $Q^*(c)$ is a decreasing function of c , we must have $\underline{c} < c_k < \bar{c}$. Also note that since $c_1 = \inf\{c | Q(c) = K\}$ and $W_Q(Q(c_1), c_1, K) > 0$ by (A4a) we also have $c_1 < c_k$. A similar argument applied to c_2 establishes $c_2 > c_k$. It is also apparent that $c_1 > \underline{c}$: for if $c_1 = \underline{c}$ we would have $Q(\underline{c}) = K$ implying that $Q(\underline{c}) = Q^*(\underline{c}) = K$ by (A4a), which violates (M). A similar argument also establishes that $c_2 < \bar{c}$. Combining our results yields

$$\underline{c} < c_1 < c_k < c_2 < \bar{c}.$$

The fact that $Q(c) = K$ for $c \in [c_1, c_2]$ follows from part (i) of Lemma 1.

To establish property (iii) of the Proposition, notice that for $c \in [c_1, c_2]$, $\pi(c) = 0$ by construction. From (A1), $\pi(c) = \int_{\underline{c}}^{c_1} [Q(z) - K] dz > 0$ for $c \in [\underline{c}, c_1]$, since $Q(c) > K$ when $c < c_1$. Similarly, $\pi(c) = \int_{c_2}^{\bar{c}} [K - Q(z)] dz > 0$ for $c \in (c_2, \bar{c}]$, since $Q(c) < K$ when $c > c_2$.

To establish property (vi) of Proposition 1, notice that (A4) implies $W_Q(Q(c), c, K) = 0$ for $c = \underline{c}, \bar{c}$. Further, $W_Q(K, c_k, K) = 0$. Thus $Q(c) = Q^*(c)$ when $c = \underline{c}, \bar{c}, c_k$.

To establish part (iv), notice that (A4a) implies $Q(c) < Q^*(c)$ when $c \in (\underline{c}, c_1)$. Recall that $Q(c) = K = Q^*(c_k)$ for all $c \in [c_1, c_2]$. Since $Q^*(c)$ is strictly decreasing in c , this implies $Q(c) = K > Q^*(c)$ for $c \in (c_1, c_k)$. Hence we have proved that $Q(c) < Q^*(c)$ for $c \in (\underline{c}, c_k)$. A similar argument suffices to establish part (v).

To complete the proof of Proposition 1 we must show that property (M) holds, which will also establish the validity of property (ii) of the Proposition. Differ-

entiating $H(c_1, c_2, K)$ with respect to K , and employing the Envelope Theorem, one can show that the necessary condition for maximization with respect to K is given by

$$(A6) \quad \int_{c_1}^{c_2} W_Q(K, c, K) f(c) dc + \int_{\underline{c}}^{c_1} F(c) dc - \int_{c_2}^{\bar{c}} (1 - F(c)) dc = 0.$$

(A6) implies (M). To see this, assume that $K \leq Q^*(\bar{c})$, for example. This implies that (IR) would bind over an interval, $[c_1, \bar{c}]$, with $W_Q(K, c, K) > 0$ for all $c \in [c_1, \bar{c}]$, which leads to a violation of (A6). A similar argument establishes that $K \geq Q^*(\underline{c})$ is also not possible. Hence (M) and part (ii) of Proposition 1 must hold.

REFERENCES

- Baron, David and Besanko, David**, "Commitment and Fairness in a Dynamic Regulatory Relationship," *Review of Economic Studies*, July 1987, 54, 413-36.
- _____ and **Myerson, Roger**, "Regulating a Monopolist with Unknown Costs," *Econometrica*, July 1982, 50, 911-33.
- Cramton, Peter, Gibbons, Robert and Klemperer, Paul**, "Dissolving a Partnership Efficiently," *Econometrica*, May 1987, 55, 615-32.
- _____ and **Palfrey, Thomas**, "Cartel Enforcement with Uncertainty about Costs," Yale University Discussion Paper, October 1986.
- Crawford, Vincent**, "The Role of Arbitration and the Theory of Incentives," in Alvin Roth, ed., *Game Theoretic Models of Bargaining*, London: Cambridge University Press, 1985.
- Feenstra, Robert and Lewis, Tracy**, "Negotiated Trade Restrictions with Private Political Pressure," University of California-Davis, Working Paper, March, 1987.
- Groves, Theodore and Ledyard, John**, "Incentive Compatibility Ten Years Later," in T. Groves, R. Radner, and S. Reiter, eds., *Information and Economic Incentives: Essays in Honor of Leonid Hurwicz*, Minneapolis: University of Minnesota Press, 1985.
- Guesnerie, Roger and Laffont, Jean-Jacques**, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics*, July 1984, 25, 329-69.
- Laffont, Jean-Jacques and Tirole, Jean**, "The Dynamics of Incentive Contracts," Massachusetts Institute of Technology, Working Paper, October 1986.
- Lewis, Tracy and Sappington, David**, "Counterbalancing Incentives in Agency Problems," University of California-Davis, Working Paper, Revised October 1988, forthcoming in *Econometrica*.
- Mirrlees, James**, "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, April 1971, 38, 175-208.
- Moore, John**, "Optimal Labour Contracts When Workers Have a Variety of Privately Observed Reservation Wages," *Review of Economic Studies*, January 1985, 52, 37-68.
- Myerson, Roger**, "Incentive Compatibility and the Bargaining Problem," *Econometrica*, January 1979, 46, 61-74.
- Riordan, Michael and Sappington, David**, "Information, Incentives, and Organizational Mode," *Quarterly Journal of Economics*, May 1987, 102, 243-64.
- Stiglitz, Joseph and Weiss, Andrew**, "Credit Rationing in Markets with Asymmetric Information," *American Economic Review*, June 1981, 71, 393-410.
- Tirole, Jean**, "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, and Organization*, Fall 1986, 2, 181-214.
- Williamson, Oliver**, *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: Free Press, 1975.