Relational Incentive Contracts

By JONATHAN LEVIN*

Standard incentive theory models provide a rich framework for studying informational problems but assume that contracts can be perfectly enforced. This paper studies the design of self-enforced relational contracts. I show that optimal contracts often can take a simple stationary form, but that self-enforcement restricts promised compensation and affects incentive provision. With hidden information, it may be optimal for an agent to supply the same inefficient effort regardless of cost conditions. With moral hazard, optimal contracts involve just two levels of compensation. This is true even if performance measures are subjective, in which case optimal contracts terminate following poor performance. (*JEL* C73, D82, L14)

Incentive problems arise in many economic relationships. Contracts that tie compensation to performance can mitigate incentive problems, but writing completely effective contracts is often impractical.[1] As a result, real-world incentives frequently are informal. Within firms, compensation and promotion often are based on difficult to verify aspects of performance such as teamwork, leadership, or initiative. Employees understand this without the precise details being codified in a formal contract and firms live up to their promises because they care about their labor market reputation.[2] Similarly, firms often expect a level of flexibility and adaptation from suppliers that goes well beyond contractual requirements. The give and take of the relationship allows prices or details of delivery to adjust in response to specific circumstances.

The need for a relational contract is a matter of degree. Sovereign nations comply with trade agreements and repay foreign debt because they desire the continued goodwill of trading partners. Politicians have an incentive to assist large donors because they will need to raise money in future campaigns. On the other hand, when a firm contracts with its employees or other firms, or when a government agency regulates an industry, a formal contract may provide a reasonable starting point. In these cases, good faith allows for more flexibility and nuance in incorporating information.

Information plays the same role in relational contracting as in standard incentive theory. Better performance measures generate more effective incentives. But a relational contract can incorporate a much broader range of subjective information. For instance, the best gauge of employee performance is often the subjective evaluations of peers or supervisors. Firms regularly use such measures for compensation and promotion decisions. In a survey of law firm compensation by the consulting firm Altman Weil, Inc., 50 percent of law firms report using

* Department of Economics, Stanford University, Stanford, CA 94305 (e-mail: jdlevin@stanford.edu). This paper is a revised chapter of my M.I.T. Ph.D. thesis. I am indebted to Glenn Ellison, Robert Gibbons, Jerry Hausman, and Bengt Holmstrom for their generous advice and encouragement. The paper has benefited from the suggestions of Susan Athey, Drew Fudenberg, Eric Maskin, Richard Levin, Paul Oyer, Antonio Rangel, Lars Stole, Steve Tadelis, Jean Tirole, and the referees. I thank the NSF for financial support.

[1] For example, it may not be practical to precisely capture what is desired in a written contract. It may be difficult even to explain the details of performance to an outsider such as the court. In some cases, the best measure of performance may be a subjective judgment. In other cases, shortcomings in the court system itself may prevent effective contractual enforcement.

[2] As an example of failed relational contracting, consider the case of United Airlines in the summer of 2001. During contract negotiations, the pilots' union urged its members to work "to the letter of our agreement." What followed was a summer of delays and cancelled flights that cost United $700 million. United's management contemplated going to court, but determined that it could not afford to further antagonize the pilots. Ultimately, negotiations resulted in a large salary increase (Roger Lowenstein, 2002).

subjective measures of performance to determine partner compensation (Altman Weil, 2000). Subjective measures are also commonly used by Wall Street firms to help determine the year-end discretionary bonuses that constitute a significant fraction of compensation.

The combination of subjectivity and discretion can give rise to costly differences of opinion. James B. Stewart (1993) describes the turmoil at First Boston in the early 1990's after year-end bonuses were lower than expected. Senior managers were incensed at what they perceived as a breach of trust. Meanwhile Credit Suisse, First Boston's parent company, argued that the bonuses merely reflected disappointing financial results. The dispute ended with many top managers leaving the firm. Lisa Endlich (1999) details a related wave of departures that occurred when Goldman Sachs dramatically cut year-end bonuses in 1994. Employees at all levels, who felt they had been promised more, quit despite the fact that senior management was predicting rapid growth.

In this paper, I provide a simple analysis of optimal relational contracting in a broad range of settings. This includes settings where there are problems of moral hazard or hidden information as well as environments where performance evaluation is purely subjective. I first show that to characterize optimal contracts it often suffices to focus on a simple class of *stationary contracts* in which the same compensation scheme is used at every date. Using this result, I explain how self-enforcement limits the scope of incentive provision and characterize optimal incentive schemes. I also show how the use of subjective performance measures necessarily gives rises to costly disputes, although optimal incentive contracts are still relatively simple.

The results are established in a general and broadly applicable agency setting. There are a principal and agent who can trade repeatedly. They can contractually specify some fixed payment and can incorporate further performance measures in a relational contract that specifies how payments and future trade will adjust to reflect performance. In the case of employment, the fixed compensation can be interpreted as a salary, while the relational contract describes how discretionary bonuses, raises, and promotions will be awarded based on performance.

In principle a relational contract can be quite complicated because performance in any given period can affect the whole future course of the relationship. Nevertheless, I show that to characterize Pareto-optimal contracts in settings of moral hazard or hidden information it suffices to consider only *stationary contracts* in which the principal promises the same contingent compensation in each period. This can be understood as follows. Even with a complex agreement, the parties have only two instruments to provide incentives—present compensation and continuation payoffs. These instruments are substitutable given risk neutrality. Thus an incentive scheme that uses continuation payoffs to reward or punish an agent can be replicated by a scheme that uses only immediate compensation. The one caveat is that direct compensation cannot replicate an outcome that punishes both parties. However, I show that when the principal's behavior is perfectly observed, optimal contracts never involve joint punishments in equilibrium. It follows that stationary contracts are optimal.

In economic terms, this result says that relationships should optimally respond to variable outcomes through adjustments in price rather than through changes in the underlying incentives. In practice, a pricing system that grants discounts to customers whose last delivery was flawed, or a compensation system that bases raises or bonuses on a review of the previous year's performance, can serve precisely this purpose.

The simple structure of stationary contracts makes it easy to characterize optimal relational incentive schemes. These schemes differ from standard incentive theory because contracts are self-enforced. Under a relational contract, each party has an incentive to pay promised compensation because reneging would bias future trade terms against the deviator or even end the relationship. But because each party has the option to walk away, the gap between the highest and lowest payments promised as a function of performance cannot exceed the present value of the relationship. This constraint limits what can be achieved in terms of incentives.

As an example, suppose that the agent has private information each period about his cost conditions. A well-known result is that optimal screening contracts will perfectly separate cost

types under certain regularity conditions. Moreover, such contracts never distort the choices of the most efficient agents. In contrast, when self-enforcement is a binding constraint, an optimal relational contract never fully distinguishes between different cost conditions and never asks for efficient effort. Indeed, the agent may be asked to provide the same (inefficient) performance without any regard for his cost conditions. These distortions arise because self-enforcement creates a shadow cost to providing effective incentives by limiting the variability of compensation.

The model predicts that relationships will be more flexible in adapting to changing cost conditions when trading is more frequent or when parties rely more on each other. The literature on vertical supply contracting suggests that adaptability is a key feature of successful long-term relationships (Oliver E. Williamson, 1985). Here, if a relationship does not generate a great deal of surplus, the terms of the relational contract will be quite rigid. In contrast, if the relationship is more productive, there will be scope to fine-tune performance to current conditions.

The use of subjective performance measures presents problems that do not arise when there is asymmetric information about costs or when performance measures are simply noisy. The reason is that a successful contract must simultaneously give the agent an incentive to perform and the principal an incentive to assess performance honestly. This makes stationary contracts ineffective. Nevertheless, if the principal reviews performance each period, an optimal contract still has a straightforward structure. The principal pays the agent a base salary each period, adding a fixed bonus if she judges performance to be above some threshold. If the principal claims nonperformance and withholds the bonus, a dispute results and the agent walks away. The optimal contract displays both information compression—the reduction of a potentially nuanced signal into just two pay levels—and conflict. Interestingly, conflict of this sort does not arise in earlier relational contracting models where parties can agree on performance measures. In contrast, the combination of subjectivity and discretion provides an explanation for the real-world compensation disputes described above.

Related Literature.—Many authors, and not just economists, have emphasized the role of ongoing relationships in contracting [early references include Stewart Macaulay (1963) and Ian R. Macneil (1974)]. In contrast to the present paper, previous relational contracting models have generally focused on environments where the parties have symmetric information [for instance, Benjamin Klein and Keith Leffler (1981); Carl Shapiro and Joseph E. Stiglitz (1984); Clive Bull (1987); and David M. Kreps (1990)]. W. Bentley MacLeod and James M. Malcomson (1989) provide a definitive treatment of this model. They show that optimal contracts can take a variety of forms: from efficiency wages where the principal pays a high fixed salary and threatens to terminate an agent for poor performance, to a system of performance bonuses paid by the principal in each period (see also MacLeod and Malcomson, 1998).

The assumption of symmetric information contrasts with the traditional incentive theory view that asymmetric information, rather than enforcement, is the central impediment to effective contracting (e.g., Bengt Holmstrom, 1979; Jean-Jacques Laffont and Jean Tirole, 1993). This paper brings these views together. I show that the connection is in fact quite close. In essence, with risk neutrality, optimal relational contracts are distinguished from standard optimal incentive contracts by the presence of a single enforcement constraint that limits the range of contingent payments.[3] This is true even if there is hidden information, moral hazard, or subjective performance evaluation.

In practice, real-world contracting often involves both formal and relational incentives. For instance, a procurement contract may include explicit penalties for late delivery, while the prospect of getting repeat business supplies the incentive to be flexible in making (nonverifiable) adaptations. George Baker et al. (1994) develop an agency model in which the principal has available both a

[3] A dynamic contracting literature in macroeconomics includes some papers that model contractual self-enforcement. Early examples include Jonathan Thomas and Tim Worrall (1988) and Andrew Atkeson (1991). This line of research focuses on the insurance and consumption-smoothing aspects of long-term relationships, problems that are absent in the present risk-neutral setup.

verifiable and a nonverifiable performance measure. They show that the principal can benefit from using both measures, but do not characterize fully optimal agreements.[4] Here I focus only on nonverifiable performance measures, but many of the results on optimal contracting extend to environments where richer formal contracts are possible (see Section II for discussion).

B. Douglas Bernheim and Michael D. Whinston (1998) develop the theme of formal and informal incentives further, arguing that when some performance measures are not verifiable, parties may deliberately leave formal contracts incomplete. This "strategic ambiguity" allows a contract to be adapted in the face of observable but nonverifiable information. Effective contracts in this paper have precisely the strategic ambiguity described by Bernheim and Whinston—the parties specify a fixed payment but leave open the possibility of adjusting this payment *ex post.* This flexibility is crucial because it allows good performance to be rewarded.

The basic problem I address—how to structure contractual relationships when the parties have useful information that cannot be effectively translated into an enforceable formal contract—also bears some relationship to analyses of the hold-up problem. Hold-ups arise when one party to a relationship makes a specific investment that is not contractible. Without contractual protection, the investing party may be unable to appropriate the returns to his investment. There are many potential responses to this problem. Sanford Grossman and Oliver Hart (1986) argue that vertical integration can substitute for contractual protection, while Aaron S. Edlin and Stefan Reichelstein (1996) show that standard legal remedies such as specific performance and expectation damages may suffice to encourage investment.

These remedies to noncontractibility do not depend on repeated interaction, but rather on the ability of the parties to renegotiate the terms of trade after performance is observed but before trade occurs. The parties structure their initial agreement to frame this renegotiation. In contrast, I focus on problems where the benefit of the agent's action accrues directly to the principal. By the time performance can be assessed, there is nothing left to negotiate over. The present model naturally describes situations where an agent performs an ongoing service for the principal (such as an employment relationship), or provides a product whose quality can only be observed in the process of using it (such as the advice provided by a consulting firm). On the other hand, hold-ups often are associated with large discrete investments of the sort that might take place at the start of a relationship.[5]

Finally, my analysis of the dynamic structure of relational contracts makes use of the repeated game techniques developed by Dilip Abreu et al. (1990) and Drew Fudenberg et al. (1994). The crucial distinction between relational contracting and the models studied in those papers is that here the parties can use monetary transfers as well as changes in equilibrium behavior to provide incentives. I discuss this point in more detail in Section II.

I. The Model

A. *The Agency Problem*

I consider two risk-neutral parties, a principal and an agent, who have the opportunity to trade at dates $t = 0, 1, 2, \ldots$. At the beginning of date t, the principal proposes a compensation package to the agent. Compensation consists of a fixed salary w_t and a contingent payment $b_t : \Phi \uparrow \mathbb{R}$, where Φ is the set of observed performance outcomes, to be described momentarily. The agent chooses whether or not to accept the principal's offer. Let $d_t \in \{0, 1\}$ denote the agent's decision.[6]

If the agent accepts, he observes a cost parameter θ_t, representing variable aspects of the environment such as task difficulty, the cost of materials, or the opportunity cost of time. The cost parameter is an independent draw from a distribution $P(\cdot)$ with support $\Theta = [\underline{\theta}, \bar{\theta}]$. The agent then chooses an action

[4] See also Klaus Schmidt and Monika Schnitzer (1995), David G. Pearce and Ennio Stacchetti (1998), and Luis Rayo (2002).

[5] Baker et al. (2002) integrate relational contracting with the hold-up literature by developing a model of repeated specific investments.

[6] Allowing the agent to propose compensation, and the principal to accept or reject, leads to an equivalent analysis.

$e_t \in \mathcal{E} \subset [0, \bar{e}]$, incurring a cost $c(e_t, \theta_t)$. The agent's cost is increasing in e with $c(e = 0, \theta) = 0$ for all θ. The agent's action generates a stochastic benefit y_t for the principal, where y_t has distribution $F(\cdot \mid e)$ and support $\mathcal{Y} = [\underline{y}, \bar{y}]$. Then $S(e, \theta) \equiv \mathbb{E}_y[y|e] - c(e, \theta)$ denotes the expected joint surplus as a function of θ and e.

During the course of trade, the agent observes all of the relevant information: his cost parameter θ_t, his action e_t and the principal's benefit y_t. The principal observes her benefit y_t, but may or may not observe θ_t and e_t. In this way, the model allows for problems of hidden information (θ is private) and moral hazard (e is private) as well as symmetric information. The performance outcome is the subset $\varphi_t \subseteq \{\theta_t, e_t, y_t\}$ observed by both parties. Let Φ be the set of possible realizations of φ_t.[7]

At the end of date t, the principal is obligated to pay the fixed salary w_t. The parties then choose whether to adjust this by the contingent payment $b_t(\varphi_t)$. This decision belongs to the principal if $b_t > 0$ and to the agent if $b_t < 0$. Let W_t denote the total payment actually made from principal to agent—either $w_t + b_t(\varphi_t)$ if the agreed payment is made, or w_t if not. Thus the agent's payoff is $W_t - c(e_t, \theta_t)$ while the principal's is $y_t - W_t$.

If the agent rejects the principal's original offer, both parties receive fixed payoffs: $\bar{u}$ for the agent and $\bar{\pi}$ for the principal. I assume these outside opportunities are less desirable than efficient trade, but sufficiently attractive that the parties weakly prefer not to trade if the agent cannot be given incentives to perform. Defining $\bar{s} \equiv \bar{u} + \bar{\pi}$, I assume that for all θ, $\max_e S(e, \theta) > \bar{s} \geq S(0, \theta)$.

Over the course of repeated interaction, the parties care about their discounted payoff stream. Starting from date t, the respective payoffs for principal and agent, discounted by a common discount factor $\delta \in [0, 1)$, are:

$$\pi_t = (1 - \delta)\mathbb{E} \sum_{\tau = t}^{\infty} \delta^{\tau - t}\{d_\tau(y_\tau - W_\tau) + (1 - d_\tau)\bar{\pi}\},$$

$$u_t = (1 - \delta)\mathbb{E} \sum_{\tau = t}^{\infty} \delta^{\tau - t}\{d_\tau(W_\tau - c(e_\tau, \theta_\tau)) + (1 - d_\tau)\bar{u}\}.$$

Here, I multiply through by $1 - \delta$ to express payoffs as a per-period average.

B. *Relational Contracts*

Because contingent compensation is merely promised, there is a temptation to renege on payments once production occurs. In a one-time interaction this means that the principal can credibly promise only the fixed salary w. As this provides no performance incentive, mutually beneficial trade cannot occur in static equilibrium and both parties receive their outside options. In contrast, ongoing interaction allows the parties to base future terms of trade on the success of present trade—potentially allowing for a good faith agreement that provides effective incentives.

A *relational contract* is a complete plan for the relationship. Let $\mathbf{h}^t = (w_0, d_0, \varphi_0, W_0, \ldots, w_{t-1}, d_{t-1}, \varphi_{t-1}, W_{t-1})$ denote the history up to date t, and $\mathcal{H}^t$ the set of possible date t histories. Then for each date t and every history $\mathbf{h}^t \in \mathcal{H}^t$, a relational contract describes: (i) the compensation the principal should offer (and which should be paid); (ii) whether the agent should accept or reject the offer; and in the event of acceptance, (iii) the action the agent should take as a function of his realized costs θ_t. Such a contract is *self-enforcing* if it describes a perfect public equilibrium of the repeated game.[8] Note that a relational contract describes behavior both on the equilibrium path and off—for instance after a party reneges on a payment.

[7] If *both* θ_t and e_t are private to the agent, the parties can generally benefit from having the agent announce θ_t prior to choosing e_t and using this announcement to tailor incentives. I ignore this possibility as it does not benefit the parties in the main cases of interest.

[8] Perfect public equilibrium requires that play following each history be a Nash equilibrium. Here, it essentially imposes the same sequential rationality requirement that subgame perfection would impose in a complete information model (Fudenberg et al., 1994).

In general, there may be many self-enforcing contracts, leading to a question of which to select. In an important sense, however, the problem of efficient contracting can be separated from the problem of distribution: if there is a self-enforcing contract that generates some surplus beyond the outside surplus $\bar{s}$, it can be divided in any way that respects the parties' participation constraints.

THEOREM 1: *If there is a self-enforcing contract that generates expected surplus $s > \bar{s}$, then there are self-enforcing contracts that give as expected payoffs any pair (u, π) satisfying $u \geq \bar{u}$, $\pi \geq \bar{\pi}$ and $u + \pi \leq s$.*

The reasoning behind Theorem 1 is that by changing the fixed compensation in the initial period of the contract, the parties can redistribute surplus without changing the incentives that are provided after the initial compensation is proposed and accepted.

Given this result, it is natural to focus on contracts that maximize the parties' joint surplus (subject to the constraint of being self-enforcing). I say that a self-enforced contract is *optimal* if no other self-enforcing contract generates higher expected surplus.

II. Stationary Contracts

This section shows that the problem of describing optimal relational contracts can be significantly simplified by focusing on contracts that take a *stationary* form.

Definition 1: A contract is stationary if on the equilibrium path $W_t = w + b(\varphi_t)$ and $e_t = e(\theta_t)$ at every date t, for some $w \in \mathbb{R}$, $b : \Phi \uparrow \mathbb{R}$, and $e : \Theta \uparrow \mathcal{E}$.

Under a stationary contract, the principal always offers the same compensation plan $w \in \mathbb{R}$ and $b : \Phi \uparrow \mathbb{R}$, and the agent always acts according to the same rule $e : \Theta \uparrow \mathcal{E}$. The contract must also specify what happens if a party fails to make a specified payment or if the principal fails to offer the expected compensation. Since these events never occur in equilibrium, there is no loss in assuming that the parties respond by breaking off trade, as this is the worst equilibrium outcome (Abreu, 1988).[9]

THEOREM 2: *If an optimal contract exists, there are stationary contracts that are optimal.*

Theorem 2 says that instead of conditioning future trade on the agent's performance, the parties can "settle up" each period using discretionary payments and then proceed to the same optimal agreement at the next date.

To understand the argument behind Theorem 2, consider a simple example where the agent has constant cost conditions and chooses an effort that results in either Low or High output. To induce the agent to exert effort, an agreement must reward the agent following High output. This reward could come in one of two forms: through an immediate payment or from moving to a continuation equilibrium that gives a high expected payoff. Given risk neutrality, the agent views these as equivalent. Thus, for any contract that relies on changes in the equilibrium behavior to provide incentives, it should be possible to provide the same incentives with payments alone. The potential catch is that discretionary payments can redistribute surplus but cannot replicate the creation or destruction of surplus that might occur if, for instance, the contract called for the agent to be fired following Low output. Because the principal's behavior is observed perfectly, however, an optimal contract would never call for surplus to be destroyed on the equilibrium path. Thus stationary contracts suffice for optimality.[10]

One small point pertains to the issue of renegotiation. The proof of Theorem 2 shows that optimal stationary contracts can be constructed to split the surplus arbitrarily provided the participation constraints are satisfied. This means that even in the event of a deviation there is no

[9] More specifically, one can assume the parties revert to statically optimizing behavior. This means that if the principal deviates by offering $(w', b') \neq (w, b)$ with $w' > \bar{u}$, the agent will accept the offer but exert zero effort.

[10] In contrast a contract that generates less than the optimal surplus might call for variation in the joint surplus over time. By placing some additional structure on the model, however, Theorem 2 can be strengthened to say that a stationary contract can achieve any surplus between $\bar{s}$ and the optimal surplus.

need to terminate the relationship. Instead, the terms of trade can be altered to hold the deviating party to his or her outside option. Say that a self-enforcing contract is *strongly optimal* if given *any* history $\mathbf{h}^t \in \mathcal{H}^t$, the continuation contract from $\mathbf{h}^t$ is optimal. A strongly optimal contract has the desirable property that even off the equilibrium path the parties cannot jointly benefit from renegotiating to a new self-enforcing contract.

COROLLARY 1: *For any optimal stationary contract, there is a strongly optimal stationary contract with the same equilibrium behavior.*

The results in this section can be applied to a variety of other contracting settings. To see why this is the case, it is useful to connect with the general theory of repeated games. Recall that in a repeated game with perfect information, optimal equilibria can always be stationary because deviations are never observed in equilibrium (Abreu, 1988). This observation applies to MacLeod and Malcomson's (1989) symmetric information model of relational contracting. Here, however, the agent's behavior is not perfectly observed. Performance might appear good or bad in equilibrium and must be rewarded and punished accordingly. In the models of Edward J. Green and Robert H. Porter (1984), Abreu et al. (1990), Fudenberg et al. (1994), and Susan Athey and Kyle Bagwell (2001), it is precisely this sort of imperfect monitoring that cause optimal equilibria to be nonstationary.

Two features make stationary relational contracts optimal despite monitoring imperfections. First, the combination of quasi-linear utility and monetary transfers means discretionary transfers can substitute for variation in continuation payoffs.[11] Second, the fact that the principal's actions are perfectly observed means that these transfers can be balanced.[12] Because Theorem 2 continues to hold in any repeated game setting with these features, it can be used to characterize optimal relational contracts in environments well beyond the ones here. One example is the model considered by Baker et al. (2002) in their study of relational contracts as a solution to repeated hold-up problems.

III. Optimal Incentive Provision

The previous section showed that in a broad range of environments it is possible to study optimal relational contracts by considering only a smaller class of stationary agreements. This section uses this observation to study optimal incentive provision in response to problems of hidden information and moral hazard. These two informational problems have been the central focus of research in incentive theory and its applications to labor and credit markets, procurement and regulation, macroeconomics and international trade.

I start by observing that relational contracts have a particular limitation. Because parties have the option to renege on payments, there are bounds on the compensation that can be credibly promised. I show how these bounds affect the structure of optimal incentives and how the resulting distortions differ from those in the classic incentive theory models where contract enforcement is not a problem.

A stationary contract is described by an effort schedule $e(\theta)$ and a payment plan $W(\varphi) = w + b(\varphi)$ that the principal offers the agent in each period. Such a contract provides per-period payoffs:

$$\pi = \mathbb{E}_{\theta,y}[y - W(\varphi) \mid e = e(\theta)],$$

$$u = \mathbb{E}_{\theta,y}[W(\varphi) - c \mid e = e(\theta)].$$

The expected joint surplus depends only on the effort schedule:

$$s = \mathbb{E}_{\theta,y}[y - c \mid e(\theta)].$$

For the compensation schedule to be

[11] The underlying connection between continuation payoffs in repeated games and monetary payments in contracting problems is already highlighted, and plays a fundamental role, in Fudenberg et al.'s (1994) derivation of the Folk Theorem.

[12] This follows because optimal continuation payoffs are necessarily on the (linear) frontier of the equilibrium payoff set. Note that this type of sequential optimality is also a feature of the complete contracting agency models of Fudenberg et al. (1990) and Malcomson and Frans Spinnewyn (1988). In those models, risk aversion can create an intertemporal smoothing reason to use future utility to reward good performance.

self-enforcing, neither the principal nor the agent must have an incentive to renege. Assuming that failure to pay results in the worst possible future outcome, a self-enforcing contract must satisfy

$$\frac{\delta}{1-\delta}(\pi - \bar{\pi}) \geq \sup_{\varphi} b(\varphi)$$

so that the principal will not walk away, and

$$\frac{\delta}{1-\delta}(u - \bar{u}) \geq -\inf_{\varphi} b(\varphi)$$

so that the agent will not walk away.

A key point here is that by adjusting the fixed component of compensation, w, up or down, slack can be transferred from one payment constraint to the other. Thus what matters is the difference between the highest payment specified under the contract and the lowest payment. This observation allows a characterization of the set of self-enforcing stationary contracts.

THEOREM 3: *An effort schedule $e(\theta)$ that generates expected surplus s can be implemented with a stationary contract if and only if there is a payment schedule $W: \Phi \uparrow \mathbb{R}$ satisfying:*

$$(IC) \quad e(\theta) \in \arg\max_{e} \mathbb{E}_y[W(\varphi)|e] - c(e, \theta)$$

for all $\theta \in \Theta$, and

$$(DE) \quad \frac{\delta}{1-\delta}(s - \bar{s}) \geq \sup_{\varphi} W(\varphi) - \inf_{\varphi} W(\varphi).$$

The dynamic enforcement (DE) constraint states that the variation in contingent payments is limited by the future gains from the relationship. The tightness of this restriction depends on the discount factor δ, as well as on the productivity of the relationship relative to the outside option $\bar{s}$. If δ is near one, the range of payments is essentially unbounded.[13] On the other hand, if δ is near zero, the range of payments may be so limited that incentives cannot be provided. The interesting question that I now address is what happens when self-enforcement is a binding constraint, but trade is nevertheless possible.

A. *Hidden Information*

Problems of hidden information arise when the principal cannot observe the precise details of the agent's environment. A classic example is a regulatory agency that can observe the quality and quantity of the delivered product, but not the marginal costs of the regulated firm. Efficiency dictates lowering price and increasing production when costs are low, but regulated firms always have an incentive to claim high costs. Similar problems arise in procurement if production costs are unknown to the procuring firm. I capture this by assuming the principal can observe the agent's behavior e_t but not his cost parameter θ_t.

An optimal relational contract displays two striking features when enforcement is a binding constraint. First, it distorts production away from the efficient level regardless of realized cost conditions. Second, even under the regularity conditions that ensure full separation of cost types in standard incentive theory models, an optimal relational contract does not fully distinguish between different cost conditions. Indeed, if enforcement is very limited, it requires the same production level regardless of cost conditions.

To proceed, I place some further structure on the model. I take both total and marginal costs to be increasing in θ and make standard assumptions about the production process. I assume that effort choice is continuous, $\mathcal{E} = [0, \bar{e}]$, and that the cost function c is differentiable with c_e, $c_{ee} > 0$, $c_\theta \geq 0$, $c_{e\theta} > 0$, and $c_{\theta ee}$, $c_{\theta e\theta} \geq 0$. I also assume that for each θ, $S(e, \theta)$ is differentiable and concave in e with an interior maximizer $e^{FB}(\theta)$. Finally I assume the distribution of costs P is concave and admits a density p. These conditions ensure that the optimal contracting problem is concave and would also ensure full separation of cost types (i.e., no pooling) in a standard self-selection setting.[14]

[13] As a consequence, the principal can induce first-best effort. As the agent is risk neutral, there may be many ways to do this, but one way is to pay the agent the full benefit y minus some constant.

[14] The usual regularity assumption is that log P is concave (i.e., that P/p is nondecreasing), which is implied by

The first step to characterizing optimal incentives is to specialize Theorem 3 to the hidden information context.

THEOREM 4: *With hidden information, an effort schedule $e(\theta)$ that generates expected surplus s can be implemented by a stationary contract if and only if (i) $e(\theta)$ is nonincreasing and (ii)*

$$(IC\text{–}DE) \qquad \frac{\delta}{1-\delta}(s-\bar{s}) \geq c(e(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} c_\theta(e(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta}.$$

Theorem 4 uses self-selection arguments to combine the (*IC*) and (*DE*) constraints from Theorem 3. The basic idea is that the payment schedule $W : [0, \bar{e}] \uparrow \mathbb{R}$ must perform two tasks. First, it must induce each type θ to select the designated effort $e(\theta)$ rather than mimicking some other type $\hat{\theta} \neq \theta$ and choosing $e(\hat{\theta})$ (similar to standard incentive compatibility). Second, it must deter all types from choosing an effort $\hat{e}$ that is not on the schedule (similar to standard interim participation).

Of these latter deviations, none is more attractive than choosing zero effort. Thus take $W(0)$ to be the lowest payment. The compensation schedule is then pinned down by self-selection: for all θ,

$$(1) \quad W(e(\theta)) = W(0) + c(e(\theta), \theta) + \int_{\underline{\theta}}^{\theta} c_\theta(e(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta}.$$

The final term is the information rent to induce self-selection. The highest transfer is $W(e(\underline{\theta}))$ and dynamic enforcement implies that $W(e(\underline{\theta})) - W(0)$ must be less than the future gains from the relationship. This gives Theorem 4.

the assumption that P is concave. This assumption and the assumption that $e^{FB}(\theta)$ is interior are easily relaxed. See Appendix B for a discussion.

The optimal contract solves:

$$\max_{e(\theta)} s = \int_{\underline{\theta}}^{\bar{\theta}} \{\mathbb{E}[y|e(\theta)] - c(e(\theta), \theta)\}\, dP(\theta)$$

$$\text{subject to } \frac{\delta}{1-\delta}(s-\bar{s}) \geq c(e(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} c_\theta(e(\theta), \theta)\, d\theta,$$

$$\text{and } e(\theta) \text{ is nonincreasing.}$$

The standard approach to solving self-selection problems is to ignore the monotonicity constraint [that $e(\theta)$ is nonincreasing] and solve the relaxed problem by pointwise optimization. It turns out, however, that the monotonicity constraint always binds so that approach cannot be used.

Instead I adopt a control theory approach. Define $\gamma(\theta) \equiv \dot{e}(\theta)$ as the control variable, and let μ and $\nu(\theta)$ denote the multipliers on, respectively, the (*IC–DE*) and monotonicity constraints. The Appendix shows that a necessary and sufficient condition for $e(\theta)$ [along with $\gamma(\theta)$, μ, $\nu(\theta)$] to solve the optimal contract problem is that:

$$(2) \quad S_e(e(\theta), \theta)p(\theta) = \frac{\mu}{1+\mu\frac{\delta}{1-\delta}}\left\{c_{e\theta}(e(\theta), \theta) - \frac{1}{\mu}\dot{\nu}(\theta)\right\}.$$

At the optimum, the marginal benefit of increasing the agent's effort in the event of each cost realization θ is equated with a shadow cost arising from the (*IC–DE*) and monotonicity constraints. The solution must also satisfy the complementary slackness conditions:

$$(3) \quad \nu(\theta) \geq 0,\ \gamma(\theta) \leq 0 \text{ and } \nu(\theta)\gamma(\theta) = 0,$$

and two boundary conditions:

$$(4) \quad \nu(\underline{\theta}) = \mu c_e(e(\underline{\theta}), \underline{\theta}) \quad \text{and} \quad \nu(\bar{\theta}) = 0.$$

The form of the optimal schedule depends on the degree to which self-enforcement is a binding constraint. At one extreme, if δ is small or outside opportunities appealing, no schedule may satisfy the constraints making production under a relational contract impossible. At the other extreme, the (*IC–DE*) constraint is slack. Then $\mu = 0$ and the first-best schedule $e^{FB}(\theta)$ is optimally chosen.

Interesting outcomes arise when production is possible but self-enforcement binds so that $\mu > 0$. In this case, the boundary conditions imply that $\nu(\underline{\theta}) > 0$, so by complementary slackness $\dot{e}(\underline{\theta}) = 0$. That is, the most efficient cost types are pooled at a single effort level. This then gives rise to two possibilities. When the enforcement constraint is very restrictive, all types are pooled at some constant effort level. When the enforcement constraint is somewhat less restrictive, the most efficient types are pooled at a relatively high effort level, while less efficient types are screened to progressively lower effort levels.

THEOREM 5: *If production is possible under a relational contract with hidden information, the optimal effort schedule* $e(\theta)$ *takes one of three forms:*

1. *Pooling:* $e(\theta)$ *is the same for all cost types.*
2. *Partial Pooling:* $e(\theta)$ *is constant on* $[\underline{\theta}, \hat{\theta}]$, *and strictly decreasing on* $(\hat{\theta}, \bar{\theta}]$, *for some* $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$.
3. *First-Best:* $e(\theta) = e^{FB}(\theta)$ *for all cost types.*

In either second-best scenario, $e(\theta) < e^{FB}(\theta)$ *for all* θ.

Surprisingly, second-best contracts call for inefficient effort regardless of cost conditions. There is a simple intuition for this. Asking for more effort from a given cost type requires an increase in the slope of the transfer schedule. Because the total variation in payments is limited, this will mean decreasing the incentives of some other cost type. Consequently, requiring an efficient level of effort for a given cost type cannot be optimal: at the margin, it would have zero benefit and a positive shadow cost.

The limit on compensation is also responsible for the pooling or partial pooling in a second-best contract. Starting from zero effort, compensation need not be increased much to raise the effort level uniformly, so it is optimal to do so. But as the effort level rises, it becomes expensive to raise the effort of less efficient types and hence optimal to screen them to lower levels of effort while the effort level of more efficient types continues to rise.

One implication is that optimal contracts are more flexible when the surplus from a relationship is greater or the interaction more frequent. Relationships that create a small amount of surplus cannot adapt to changing cost conditions. As an illustration, imagine a supply relationship in manufacturing. Even with simple fixed-price contracts each period, a functioning relationship may allow a base level of noncontractible quality above the bare minimum. A close supply relationship can make it possible to tailor details of delivery or specific quality requirements more closely to current cost conditions, even if these conditions are known only to one party.

In the present setting, an optimal contract balances efficiency with what is effectively a limit on total incentives. In contrast, in the classical screening problem where the principal offers a profit-maximizing contract to the agent subject to an interim participation constraint, the fundamental trade-off is between efficiency and rent extraction. Increased incentives boost efficiency but mean that the principal must leave information rents to the agent. For a given type θ, this leads to a shadow cost of incentives that is proportional to $P(\theta)/p(\theta)$. Since $P(\underline{\theta}) = 0$, the most efficient type produces efficiently, but not the others. Here, Theorem 1 implies that rent extraction is not an issue; rather, the problem is self-enforcement. The different trade-offs generate quite different predictions.

Although the motivation is different, the limit on contingent compensation imposed by self-enforcement also can be interpreted as representing a form of limited liability—i.e., a requirement that $\underline{W} < W(e) < \bar{W}$ for some (here endogenous) liability limits $\underline{W}$ and $\bar{W}$. Models of contracting under limited liability date back to David Sappington (1983), who studied hidden information problems under the *one-sided* constraint that $W(e) \geq 0$.[15] Sapping-

[15] Many papers study moral hazard problems with a one-sided limited liability constraint on the agent. The idea

ton's work showed that a one-sided constraint has the same qualitative effect as the standard interim participation constraint. An implication of Theorem 5 is that hidden information incentives differ dramatically when both parties have limited resources.

B. *Moral Hazard*

With hidden information, optimal contracts can reflect a rich trade-off between screening, incentives, and enforcement. In moral hazard environments, where the principal observes only a noisy measure of performance, self-enforcement leads to a very stark form of optimal contract. Although there may be many levels of measured performance, an optimal contract compresses this information into just two levels of compensation. The principal sets a base payment that is adjusted up if y_t exceeds some threshold or down if y_t misses the threshold.

To focus on moral hazard, I now assume that the agent's cost parameter θ_t is observable but that his action e_t is not. Thus a stationary contract specifies compensation $W(\theta, y) = w + b(\theta, y)$ and an associated effort schedule $e(\theta)$. As in the previous section, I assume that effort is a continuous choice and that $c_e, c_{ee} > 0$. In addition, I assume the Mirrlees-Rogerson conditions are satisfied: that is, the density of $F(y|e)$ has the monotone likelihood ratio property and $F(y|e = c^{-1}(x; \theta))$ is convex in x for any θ.[16]

THEOREM 6: *An optimal contract implements some effort schedule* $e(\theta) \leq e^{FB}(\theta)$. *For each* θ, *either* $e(\theta) = e^{FB}(\theta)$ *or* $e(\theta) < e^{FB}(\theta)$ *and the payments* $W(\theta, y)$ *are "one-step":* $W(\theta, y)$ *equals* $\underline{W}$ *for all* $y < \hat{y}(\theta)$ *and* $\bar{W}$ *for all* $y \geq \hat{y}(\theta)$, *where* $\bar{W} = \underline{W} + \frac{\delta}{1-\delta}(s - \bar{s})$ *and* $\hat{y}(\theta)$ *is the point at which the likelihood ratio* $(f_e/f)(y|e(\theta))$ *switches from negative to positive as a function of* y.

The intuition for Theorem 6 is that the parties' risk neutrality makes it desirable to use the strongest possible incentives. Self-enforcement limits the maximal reward and punishment. The "one-step" scheme that results is reminiscent of the static model of Robert D. Innes (1990). Innes shows that under a one-sided limited liability requirement that $W(y) \geq 0$ and the further requirement that $W(y) \leq y$, the optimal contract pays 0 for low outcomes and y for high outcomes. Here self-enforcement imposes a similar lower bound and also a fixed upper bound. These bounds define the two payment levels.

IV. Subjective Performance Measures

The previous sections assumed that although performance measures might be imperfect and noncontractible, the parties at least could agree on them. In practice, relational contracting often involves performance measures that are inherently *subjective*—for instance, the opinions of supervisors or peers. Firms that rely heavily on their incentive systems, such as the Lincoln Electric Company (Norman Fast and Norman Berg, 1975), often make extensive use of subjective performance reviews. Incorporating subjective assessments allows for a more nuanced form of compensation, but also creates the potential for disputes over how to interpret past performance.

I model subjective performance measurement by assuming that the agent's action e_t is privately chosen, while the principal privately observes the output y_t. That is, the principal has a noisy signal of the agent's performance not based on an objective measure the agent can observe. Having learned y_t, the principal can deliver a report $m_t \in \mathcal{M}$ (where $\mathcal{M}$ is some large set of possible messages). Compensation at time t is composed of a base payment $w_t \in \mathbb{R}$ and an adjustment $b_t : \mathcal{M} \uparrow \mathbb{R}$. For simplicity, I suppose that the agent's environment is not stochastic, i.e., $c : [0, \bar{e}] \uparrow \mathbb{R}$ is the same in each period, and I maintain the assumptions on costs and output from the previous section.

With subjective performance evaluation, stationary contracts are no longer effective. To

is that it becomes costly (in terms of rents) for the principal to provide incentives even if the agent is risk neutral. Rosella Levaggi (1999) sets up, but does not solve out, a hidden information problem with two-sided limited liability.

[16] These conditions ensure that the agent's incentive compatibility constraint can be replaced with a first-order condition for his action choice (William P. Rogerson, 1985).

see why, observe that the principal will only be willing to make distinct reports m and m' in response to distinct outcomes y and y' if the two reports yield the same future profits [under a stationary contract, this would require that $b(m) = b(m')$]. Thus if a relational contract is to provide both productive and monitoring incentives, the agent's payoff going forward must vary with performance but the principal's must not. It follows that production cannot be optimal after every history. Joint punishments or disputes must occur in equilibrium.

The question then is how best to provide incentives for the agent while inducing the principal to make honest evaluations. This problem is complicated by the great flexibility in structuring performance reviews over time.[17] To make progress, I focus on contracts where the principal provides a full performance evaluation after each period. That is, I consider contracts with the following *full review* property: given any history up to t and compensation offer at t, then any two outputs $y_t \neq y_t'$ must generate distinct reports $m_t \neq m_t'$.[18] Given this restriction, it is natural to suppose that these contracts simply call for the principal to report her true assessment at each date—i.e., call for her to report $m_t = y_t$.

To characterize optimal full review contracts, I first define a simple form of termination contract. Like a stationary contract, a termination contract requires the same compensation in each period that trade occurs, but it also allows for the possibility that the parties might dissolve the relationship.

Definition 2: A contract is a termination contract if in every period t that trade occurs, $W_t = w + b(y_t)$, $e_t = e$, $m_t = y_t$ and trade continues beyond t with probability $\alpha_t = \alpha(y_t)$ and otherwise ceases forever, for some $w \in \mathbb{R}$, $b : \mathcal{Y} \uparrow \mathbb{R}$, $e \in [0, \bar{e}]$, and $\alpha : \mathcal{Y} \uparrow [0, 1]$.

[17] Specifically, this involves looking for optimal equilibria in a repeated game with private monitoring, a problem that has seen little theoretical progress.

[18] The full review property means that in equilibrium, the principal does not maintain any private information from period to period. This restriction can also be stated in terms of the equilibrium concept. I focus on perfect public equilibria of the repeated game rather than sequential equilibria.

In environments with subjective performance measures, termination contracts are optimal among all contracts with the full review property. The logic is by now familiar. Rather than using varied continuation behavior to provide incentives, the parties can use a combination of immediate compensation and termination. As above, this means that rather than having to study a range of potentially complex relational contracts, it is possible to focus on a simple subset.

THEOREM 7: *If an optimal full review contract exists, a termination contract can achieve this optimum. An optimal termination contract takes a one-step form. It specifies some* $e \leq e^{FB}$ *and a threshold* $\hat{y}$*: if* $y_t < \hat{y}$, $W_t = w$ *and the relationship terminates; if* $y_t \geq \hat{y}$, $W_t = w + b$ *and the relationship continues. Compensation is given by* $w = \bar{u} + c(e) + k$ *and* $b = \frac{\delta}{1-\delta}(s - \bar{s} - k)$ *for some* $k \in [0, s - \bar{s}]$*, where the expected surplus is*

$$s = \bar{s} + \frac{(1-\delta)\mathbb{E}[y - c - \bar{s} | e]}{1 - \delta[1 - F(\hat{y}|e)]}.$$

The reason for the one-step form of optimal compensation is again that the strongest incentives come from maximal and minimal rewards. The difference here is that the principal must not have a preference for reporting a bad outcome. Thus an output $y < \hat{y}$ results in no bonus, but also in separation.[19]

The fact that even optimal contracts involve disputes creates a role for mediation or dispute resolution. Many firms use mediation systems to resolve internal disputes or disputes with suppliers. Proponents of these systems often argue that they help preserve important relation-

[19] An additional difference here is that an optimal contract cannot involve discretionary downward adjustments to the base compensation. The reason is that a poor performance review is always followed by termination so the agent would never volunteer to take reduced compensation. An optimal contract might involve discretionary bonuses, $b > 0$, or simply a flat efficiency wage where effort is motivated by the threat of being laid off. One consequence is that explicit communication (i.e., the announcement m) is unnecessary, since the principal can "communicate" simply by unilaterally raising compensation or by firing the agent.

ships (Todd Carver and Albert A. Vondra, 1994). Here, because inefficiency results from the parties having different beliefs about performance, a mediator can create value by providing objective information that narrows the space of disagreement, hence allowing a transfer that prevents the relationship from breaking up.[20]

Even apart from instituting formal mechanisms such as mediation, one might ask whether there are alternative systems for subjective performance evaluation that improve on the one studied here. It turns out in this regard that a full performance review each period is limiting in a subtle way. Some evidence of this is provided by the fact that even as $\delta \uparrow 1$ the optimal surplus under a full review contract remains bounded away from the first-best as a consequence of endogenous separations.[21] Work by Abreu et al. (1991), Olivier Compte (1998), and Michihiro Kandori and Hitoshi Matsushima (1998) on repeated games suggests a clever improvement. They suggest that the principal could review performance only every T periods. With a full review every T periods, the optimal contract takes a similar form—a bonus for a favorable review and termination following a poor review—but the threshold for termination depends on performance in all T preceding periods. For sufficiently large δ, the delay in reviewing performance can improve incentives by allowing the principal to more accurately assess performance at each review date.[22]

[20] MacLeod (2003) extends the model in this section to allow the agent to observe a signal that is correlated with the principal's assessment. He shows that higher correlation increases efficiency.

[21] See the secondary Appendix, available from the author, for a proof of this result.

[22] In fact, a T-review contract can allow an approximately first-best outcome as $\delta \uparrow 1$ if the parties also let $T \uparrow \infty$. The idea is that with T-review, the threshold for termination takes the form of a likelihood ratio test. As $T \uparrow \infty$, the power of this test becomes very high, allowing for very small termination probabilities (see the cited papers for details).

This leaves open the question of whether there are still more effective systems of ongoing performance review. The previous paragraph suggests a benefit to infrequent reviews. But in any realistic setting, there are likely to be benefits to providing frequent feedback that go beyond adjusting compensation. To take an obvious example, agents cannot learn from mistakes without being informed of them. Further research is needed to understand how performance review systems should resolve these trade-offs.

V. Conclusion

Good faith is an essential ingredient to many contracting relationships. This paper has taken the view that even in relatively complex environments where parties may not be able to precisely observe each other's costs or efforts, it may be possible to construct good faith agreements that parties will live up to. These contracts are limited relative to what could be achieved if an infallible court system could monitor and enforce all agreements. Nevertheless, I have shown that this limitation can be characterized succinctly in the form of bounds on promised compensation. As a result, optimal relational incentive schemes can be characterized and compared with predictions from standard incentive theory. For instance, I showed that optimal contracts will forgo screening for private cost information unless parties are sufficiently patient, and even then will involve systematic distortions from efficiency. I also considered the difference between objective and subjective nonverifiable measures of performance. More generally, the framework developed here may perhaps find application in many areas where incentive theory has provided fruitful insights—regulation, employment contracting, vertical supply contracting—but where the assumption that parties can perfectly commit themselves via detailed written contracts is strained.

Appendix A: Stationary Contracts

This Appendix proves Theorems 1–3 and Corollary 1. To begin, consider a contract that in its initial period calls for payments w, $b(\varphi)$, and effort $e(\theta)$. If the offer is made and accepted and the discretionary payment made, suppose the continuation contract gives payoffs $u(\varphi)$, $\pi(\varphi)$ as a

function of the observed outcome φ. A deviation—an unexpected offer or rejection or a refusal to make the discretionary payment—implies reversion to the static equilibrium. There is no loss in assuming this worst punishment off the equilibrium path of play (Abreu, 1988). Define $W(\varphi) \equiv w + b(\varphi)$.

Let u, π be the expected payoffs under this contract:

$$u \equiv (1 - \delta)\mathbb{E}_{\theta,y}[W(\varphi) - c|e(\theta)] + \delta\mathbb{E}_{\theta,y}[u(\varphi)|e(\theta)],$$

$$\pi \equiv (1 - \delta)\mathbb{E}_{\theta,y}[y - W(\varphi)|e(\theta)] + \delta\mathbb{E}_{\theta,y}[\pi(\varphi)|e(\theta)].$$

Let $s \equiv u + \pi$ denote the expected contract surplus:

$$s \equiv (1 - \delta)\mathbb{E}_{\theta,y}[y - c|e(\theta)] + \delta\mathbb{E}_{\theta,y}[s(\varphi)|e(\theta)]. \tag{A1}$$

where $s(\varphi) \equiv u(\varphi) + \pi(\varphi)$ is the continuation surplus following outcome φ.

This contract is self-enforcing if and only if the following conditions hold: (i) the parties are willing to initiate the contract:

$$u \geq \bar{u} \quad \text{and} \quad \pi \geq \bar{\pi};$$

(ii) for all θ, the agent is willing to choose $e(\theta)$,

$$e(\theta) \in \arg\max_{e} \mathbb{E}_{y}\left[W(\varphi) + \frac{\delta}{1 - \delta} u(\varphi) \middle| e \right] - c(e, \theta);$$

(iii) for all φ, both parties are willing to make the discretionary payment,

$$b(\varphi) + \frac{\delta}{1 - \delta} u(\varphi) \geq \frac{\delta}{1 - \delta} \bar{u},$$

$$-b(\varphi) + \frac{\delta}{1 - \delta}\pi(\varphi) \geq \frac{\delta}{1 - \delta}\bar{\pi};$$

and (iv) each continuation contract is self-enforcing. In particular, for each φ, the pair $u(\varphi)$, $\pi(\varphi)$ correspond to a self-enforcing contract that will be initiated in the next period.

PROOF OF THEOREM 1:

Consider changing the initial compensation w in the above contract. This changes the expected payoffs u, π but not the joint surplus s. If the original contract was self-enforcing, the altered contract will still satisfy (ii)–(iv) as these do not depend on w, and will satisfy (i) so long as the resulting payoffs u', π' satisfy $u' \geq \bar{u}$ and $\pi' \geq \bar{\pi}$. In this event, the new contract also will be self-enforcing.

Let s^* be the maximum surplus generated by any self-enforcing contract. By Theorem 1, the set of payoffs possible under a self-enforcing contract is then $\{(u, \pi) : u \geq \bar{u}, \pi \geq \bar{\pi} \text{ and } u + \pi \leq s^*\}$. Thus for a given contract described as above to be self-enforcing, its continuation payoffs for each φ must satisfy:

$$(u(\varphi), \pi(\varphi)) \in \{(u, \pi) : u \geq \bar{u}, \pi \geq \bar{\pi}, \text{ and } u + \pi \leq s^*\}.$$

In particular, for all φ, $\bar{s} \leq s(\varphi) \leq s^*$.

LEMMA 1: *Any optimal contract is sequentially optimal. Following any history that occurs with positive probability in equilibrium, it specifies an optimal continuation contract.*

PROOF OF LEMMA 1:

Consider increasing $\pi(\varphi)$ for some φ in the above contract. This does not affect the effort constraint (ii), relaxes the participation and payment constraints (i) and (iii), and is consistent with a self-enforced continuation contract, (iv), so long as $u(\varphi) + \pi(\varphi) \leq s^*$. Moreover, this change increases the expected surplus s. Thus for a contract to be optimal, it must have $u(\varphi) + \pi(\varphi) = s^*$ for all φ that occur with positive probability. As this argument also applies to each continuation contract, it follows that any optimal contract must be sequentially optimal.

PROOF OF THEOREM 2:

Suppose that the contract specified above is optimal and generates surplus $s = s^*$. By Lemma 1, $s(\varphi) = s^*$ for all φ that occur with positive probability in the initial period. Hence if $e(\theta)$ is the initial effort schedule, (A1) implies that

$$\mathbb{E}_{\theta,y}[y - c|e(\theta)] = s^*.$$

I now construct a stationary contract that implements $e(\theta)$ in every period and thus is optimal. The idea is to take the original contract and transfer any variation in the continuation payoffs $u(\varphi)$, $\pi(\varphi)$ into the discretionary payments.

Let $u^* \in [\bar{u}, s^* - \bar{\pi}]$ be given, and define stationary discretionary payments:

$$b^*(\varphi) \equiv b(\varphi) + \frac{\delta}{1-\delta} u(\varphi) - \frac{\delta}{1-\delta} u^*.$$

From here, define the fixed payment w^* so that the agent's per-period payoff will be u^*:

$$w^* \equiv u^* - \mathbb{E}_{\theta,y}[b(\varphi) - c|e(\theta)].$$

Let $W^*(\varphi) \equiv w^* + b^*(\varphi)$.

The stationary contract has the principal propose w^*, $b^*(\varphi)$ in each period and the agent respond with effort $e(\theta)$. Any deviation causes reversion to static equilibrium behavior. This contract gives the agent an expected payoff: $\mathbb{E}_{\theta,y}[W^*(\varphi) - c|e(\theta)] = u^*$ and the principal $\pi^* \equiv s^* - u^*$.

To see that this stationary contract is self-enforcing, observe that (i) by definition $u^* \geq \bar{u}$ and also $\pi^* \geq \bar{\pi}$. Moreover, the discretionary payments are defined so that for all φ,

$$b^*(\varphi) + \frac{\delta}{1-\delta} u^* = b(\varphi) + \frac{\delta}{1-\delta} u(\varphi). \tag{A2}$$

Consequently, the agent's expected future payoff at the time of choosing his effort and both parties' expected payoffs at the time of making discretionary payments are precisely the same as in the initial period of the original contract.

Substituting (A2) into (ii) and (iii) above implies that the stationary contract defined by w^*, $b^*(\varphi)$, $e(\theta)$, u^*, and π^* will satisfy (ii) and (iii). Thus the agent will choose $e(\theta)$ and there is no incentive to renege on a discretionary payment. Finally, because this stationary contract repeats in each following period, the continuation contract is self-enforcing, (iv). Thus the constructed contract is self-enforcing and gives a per-period surplus s^*.

PROOF OF COROLLARY 1:

The proof of Theorem 2 constructs a family of optimal stationary contracts giving the agent any payoff $u^* \in [\bar{u}, s^* - \bar{\pi}]$ and the principal a corresponding payoff $\pi^* = s^* - u^*$. Each of these

contracts remains optimal on the equilibrium path, but reverts to no trade following an observed deviation. Suppose that in a period following an observed deviation, the parties instead initiate either a stationary optimal contract that gives the agent $\bar{u}$ (if the agent deviated), or one that gives the principal $\bar{\pi}$ (if the principal deviated). The resulting contract is still stationary, is optimal after any history $\mathbf{h}^t$, and is self-enforcing because it provides a payoff-equivalent punishment for an observed deviation.

PROOF OF THEOREM 3:

($\Rightarrow$) Consider a self-enforcing stationary contract with effort $e(\theta)$, payments $W(\varphi) = w + b(\varphi)$, and per-period payoffs u, π. Since for any θ, the agent can choose $e \neq e(\theta)$ and continue with the contract, (*IC*) is a necessary condition for self-enforcement. Similarly, as either party can renege on a discretionary payment and exit the relationship, then for all φ:

$$\text{(A3)} \qquad \frac{\delta}{1-\delta}(u - \bar{u}) \geq -b(\varphi) \quad \text{and} \quad \frac{\delta}{1-\delta}(\pi - \bar{\pi}) \geq b(\varphi).$$

The agent's constraint must hold for $\inf_\varphi b(\varphi)$ and the principal's for $\sup_\varphi b(\varphi)$. Summing these constraints implies that

$$\frac{\delta}{1-\delta}(s - \bar{s}) \geq \sup_{\varphi} b(\varphi) - \inf_{\varphi} b(\varphi).$$

Since $b(\varphi) = W(\varphi) - w$, this is equivalent to (*DE*).

($\Leftarrow$) Conversely, suppose there is a payment schedule $W(\varphi)$ and an effort profile $e(\theta)$ that satisfy (*IC*) and (*DE*). Define:

$$b(\varphi) \equiv W(\varphi) - \inf_{\varphi} W(\varphi),$$

and a corresponding fixed payment:

$$w \equiv \bar{u} - \mathbb{E}_{\theta,y}[b(\varphi) - c | e(\theta)].$$

Now consider a stationary contract with payments w, $b(\varphi)$, and effort $e(\theta)$ and deviations punished with reversion to the static equilibrium. This contract gives per-period payoffs $\bar{u}$ to the agent and $\pi \equiv s - \bar{u}$ to the principal. By (*DE*), $s \geq \bar{s}$, so $\pi \geq \bar{\pi}$ and both parties are willing to initiate the contract. Moreover, (*IC*) implies that for any θ, the agent prefers $e(\theta)$ to any $e \neq e(\theta)$, while from (*DE*) it is easy to check that (A3) both hold, so there is no incentive to renege on payments.

Appendix B: Optimal Incentive Provision

PROOF OF THEOREM 4:

By Theorem 3, there is a stationary self-enforcing contract with effort $e(\theta)$ if and only if there exist payments $W(e)$ satisfying (*IC*) and (*DE*). I first derive necessary and sufficient conditions for $e(\theta)$, $W(e)$ to satisfy (*IC*). I then extend this to conditions under which both (*IC*) and (*DE*) can be satisfied.

Let $W(e)$ be a stationary payment schedule and define $U(\theta)$ as the resulting single-period payoff for an agent with cost type θ:

$$U(\theta) \equiv \max_{e} W(e) - c(e, \theta).$$

Regardless of the payment schedule, $U(\theta)$ will be decreasing in θ as a consequence of the Envelope Theorem.

Given that the agent selects his effort to maximize $W(e) - c(e, \theta)$, the schedule $W(e)$ will induce effort $e(\theta)$ [i.e., will satisfy (*IC*)] if and only if (i) $e(\theta)$ is nonincreasing; (ii) for all θ,

$$U(\theta) = W(e(\theta)) - c(e(\theta), \theta) = U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} c_{\theta}(e(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta}; \tag{B1}$$

and finally (iii) for all θ and $\hat{e} \notin \{e(\theta)\}_{\theta \in \Theta}$,

$$U(\theta) = W(e(\theta)) - c(e(\theta), \theta) \geq W(\hat{e}) - c(\hat{e}, \theta). \tag{B2}$$

Conditions (i) and (ii) follow from standard incentive compatibility arguments. They imply that each type θ prefers to choose $e(\theta)$ rather than mimic type θ' by choosing $e(\theta')$. Condition (iii) implies that each type θ prefers $e(\theta)$ to some $\hat{e}$ that is not a specified choice for any cost type. It is analogous to an interim participation constraint.

Now, assume $e(\theta)$ and $W(e)$ satisfy (*IC*). Then $e(\theta)$ is nonincreasing, and rewriting (B1),

$$W(e(\theta)) = c(e(\theta), \theta) + U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} c_{\theta}(e(\tilde{\theta}), \tilde{\theta})\, d\tilde{\theta}. \tag{B3}$$

Moreover, since any type can deviate to $e = 0$, and $c(0, \theta) = 0$, (iii) implies that

$$U(\bar{\theta}) = W(e(\bar{\theta})) - c(e(\bar{\theta}), \bar{\theta}) \geq W(0). \tag{B4}$$

Combining (B4) and (B3) implies that

$$W(e(\underline{\theta})) - W(0) \geq c(e(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} c_{\theta}(e(\theta), \theta)\, d\theta.$$

Thus a *necessary* condition for $e(\theta)$ and $W(e)$ to satisfy both (*IC*) and (*DE*) is that

$$\frac{\delta}{1 - \delta}(s - \bar{s}) \geq c(e(\underline{\theta}), \underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} c_{\theta}(e(\theta), \theta)\, d\theta. \tag{*IC–DE*}$$

Conversely, suppose that $e(\theta)$ is nonincreasing and satisfies (*IC–DE*). I construct a payment plan $W(e)$ that satisfies (*IC*) and (*DE*). Let $W(0)$ be given arbitrarily. Define $W(e(\bar{\theta}))$ to satisfy (B4) with equality. Then $U(\bar{\theta}) = W(e(\bar{\theta})) - c(e(\bar{\theta}), \bar{\theta}) = W(0)$. For each $\theta \in [\underline{\theta}, \bar{\theta})$, define $W(e(\theta))$ to satisfy (B3), and for each $\hat{e} \notin \{e(\theta)\}_{\theta \in \Theta}$, define $W(\hat{e}) = W(0)$.

I now verify that $e(\theta)$, $W(e)$ satisfy both (*IC*) and (*DE*). By construction, $e(\theta)$ and $W(e)$ automatically satisfy (i) and (ii). And because $U(\theta) = \max_{e} W(e) - c(e, \theta)$ is nonincreasing in θ and $c(\hat{e}, \theta) \geq 0$, the fact that $U(\bar{\theta}) = W(0)$ by definition implies that (iii) is also satisfied. Consequently, $e(\theta)$, $W(e)$ satisfy (*IC*). Now, because $e(\theta)$ is nonincreasing, the construction of W using (B3) implies that W assumes its largest value at $W(e(\underline{\theta}))$ and smallest at $W(0)$. Since (*IC–DE*) implies precisely that $\frac{\delta}{1 - \delta}(s - \bar{s}) \geq W(e(\underline{\theta})) - W(0)$, it follows that (*DE*) is satisfied, completing the proof.

Optimal Hidden Information Contracting.—I now derive the solution to the optimal contracting problem with hidden information. To formulate this as a problem in optimal control, take $\gamma(\theta) \equiv \dot{e}(\theta)$ to be the control and $e(\theta)$ to be a state variable. It is useful to define a second state variable,

$$K(\theta) \equiv \int_{\underline{\theta}}^{\theta} \left\{ \frac{\delta}{1-\delta} S(e(\tilde{\theta}), \tilde{\theta}) p(\tilde{\theta}) - c_{\theta}(e(\tilde{\theta}), \tilde{\theta}) \right\} d\tilde{\theta}.$$

The control problem is given by the Hamiltonian:

$$H = S(e(\theta), \theta)p(\theta) + \eta(\theta)\gamma(\theta) + \lambda(\theta)\left[\frac{\delta}{1-\delta} S(e(\theta), \theta)p(\theta) - c_{\theta}(e(\theta), \theta) \right],$$

with the monotonicity constraint:

$$\gamma(\theta) \leq 0$$

and the (*IC–DE*) constraint (now rewritten as a boundary constraint):

$$\text{(B5)} \qquad K(\underline{\theta}) = 0 \quad \text{and} \quad K(\bar{\theta}) - \frac{\delta}{1-\delta} \bar{s} - c(e(\underline{\theta}), \underline{\theta}) \geq 0.$$

Here $\eta(\theta)$ and $\lambda(\theta)$ are co-state variables assigned to $\dot{e}(\theta)$ and $\dot{K}(\theta)$.

This problem is concave, so the Pontryagin conditions are both necessary and sufficient for a solution (e.g., Daniel Léonard and Ngo van Long, 1992, pp. 250–51). Letting $\nu(\theta)$ be the multiplier on the monotonicity constraint and μ the multiplier on the (*IC–DE*) inequality, the solution satisfies:

$$\nu(\theta) = H_{\gamma} = \eta(\theta)$$

$$-\dot{\eta}(\theta) = H_{e} = S_{e}(e(\theta),\theta)p(\theta)\left(1 + \lambda(\theta)\frac{\delta}{1-\delta} \right) - \lambda(\theta)c_{e\theta}(e(\theta),\theta)$$

$$-\dot{\lambda}(\theta) = H_{k} = 0$$

$$\dot{e}(\theta) = H_{\eta} = \gamma(\theta)$$

$$\dot{K}(\theta) = H_{\lambda} = \frac{\delta}{1-\delta} S(e(\theta),\theta)p(\theta) - c_{\theta}(e(\theta),\theta);$$

as well as the boundary conditions:

$$K(\underline{\theta}) = 0,\ \lambda(\bar{\theta}) = \mu,\ \eta(\underline{\theta}) = \mu c_{e}(e(\underline{\theta}), \underline{\theta}), \text{ and } \eta(\bar{\theta}) = 0,$$

the complementary slackness conditions on the monotonicity constraint:

$$\nu(\theta) \geq 0, \quad \gamma(\theta) \leq 0 \quad \text{and} \quad \nu(\theta)\gamma(\theta) = 0,$$

and the self-enforcement constraint:

$$\mu \geq 0,\ K(\bar{\theta}) - \frac{\delta}{1-\delta}\bar{s} - c(e(\underline{\theta}),\ \underline{\theta}) \geq 0$$

$$\text{and} \quad \mu\left[K(\bar{\theta}) - \frac{\delta}{1-\delta}\bar{s} - c(e(\underline{\theta}),\underline{\theta})\right] = 0.$$

To obtain the conditions in the text, I make two observations. First, because $\dot{\lambda}(\theta) = 0$ for all θ, and $\lambda(\bar{\theta}) = \mu$, it must be that $\lambda(\theta) = \mu$ for all θ. In addition, because $\eta(\theta) = \nu(\theta)$ for all θ, then $\dot{\eta}(\theta) = \dot{\nu}(\theta)$ as well. Thus, it is possible to substitute for $\eta(\theta)$ and $\lambda(\theta)$ in the above equations, leading to the conditions in the text.

Beyond the conditions in the text, the solution must also satisfy the complementary slackness conditions on the self-enforcement constraint. This leads to two cases, depending on whether self-enforcement is binding.

Case 1: Suppose that $\mu = 0$. Then (*IC–DE*) is slack at the solution, so it suffices to consider the problem of maximizing the joint surplus subject to the monotonicity constraint. Because $e^{FB}(\theta)$ is decreasing, it both maximizes the joint surplus and satisfies monotonicity. Hence $e(\theta) = e^{FB}(\theta)$ at the optimum.

Case 2: Suppose that $\mu > 0$. Then (*IC–DE*) binds at the solution. I proceed in a series of steps. I first define the schedule $e^R(\theta)$ as the unique solution to:

$$S_e(e^R(\theta),\ \theta)p(\theta) = \frac{\mu}{1 + \mu\dfrac{\delta}{1-\delta}}\{c_{e\theta}(e^R(\theta),\ \theta)\}. \tag{B6}$$

Under the assumptions in the text, $e^R(\theta)$ is decreasing in θ, and $e^R(\theta) < e^{FB}(\theta)$ for all θ.

LEMMA 2: *On any interval where the solution $e(\theta)$ is decreasing, $e(\theta) = e^R(\theta)$.*

PROOF:

On any interval where $\dot{e}(\theta) < 0$, it must be that $\nu(\theta) = 0$. Hence $\dot{\nu}(\theta) = 0$ as well. Eliminating $\dot{\nu}$ in the Pontryagin condition (2) yields the result.

LEMMA 3: *If for some $\hat{\theta}$, $\dot{e}(\hat{\theta}) < 0$, then $e(\theta)$ is decreasing on $[\hat{\theta},\ \bar{\theta}]$.*

PROOF:

I argue by contradiction. Suppose that for some θ_0, $e(\theta)$ is decreasing below θ_0 and constant above it. By complementary slackness, $\nu(\theta_0^-) = 0$ and $\nu(\theta_0^+) > 0$. Hence $\dot{\nu}(\theta_0) > 0$. Moreover, it follows from the Pontryagin condition (2) that over any interval where $e(\theta)$ is constant, $\dot{\nu}(\theta)$ must be increasing (using the assumption that S_e and p are decreasing in θ and that $c_{e\theta}$ is increasing in θ). Thus, $\dot{\nu}(\theta)$ is both positive and increasing above θ_0. Consequently $\nu(\theta) > 0$ for all $\theta > \theta_0$, contradicting the boundary condition that $\nu(\bar{\theta}) = 0$.

PROOF OF THEOREM 5:

I use the two lemmas above and the optimality conditions in the text. First, observe that at the optimum $\nu(\underline{\theta}) = \mu c_e(e(\underline{\theta}),\ \underline{\theta})$. Thus, if self-enforcement is binding, $\nu(\underline{\theta}) > 0$, which implies by complementary slackness that $\dot{e}(\underline{\theta}) = 0$. So there must be pooling of the most efficient types. Combined with the two lemmas, this yields two possibilities: either $e(\theta)$ is constant below some $\hat{\theta} \in (\underline{\theta},\ \bar{\theta})$ and decreasing above it, or $e(\theta)$ is constant on the entire interval $[\underline{\theta},\ \bar{\theta}]$.

If there is partial pooling, then for all $\theta \geq \hat{\theta}$, $e(\theta) = e^R(\theta)$, while for all $\theta \leq \hat{\theta}$, all types are pooled at some $\hat{e}$. By continuity, this means that $\hat{e} = e^R(\hat{\theta})$. To identify the "cut-off" type $\hat{\theta}$, observe that $v(\underline{\theta}) = \mu c_e(e(\underline{\theta}), \underline{\theta})$ and $v(\hat{\theta}) = 0$. Integrating the Pontryagin condition (2) from $\underline{\theta}$ up to $\hat{\theta}$, and substituting these boundary conditions yields:

$$\text{(B7)} \qquad \int_{\underline{\theta}}^{\hat{\theta}} S_e(e^R(\hat{\theta}), \theta)p(\theta)\, d\theta = \frac{\mu}{1 + \mu \dfrac{\delta}{1 - \delta}} \{c_e(e^R(\hat{\theta}), \hat{\theta})\}.$$

From the assumptions on primitives, it is easy to check that there is at most one value $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ that satisfies (B7). If this $\hat{\theta} < \bar{\theta}$, the solution has $e(\theta) = e^R(\hat{\theta})$ for $\theta \leq \hat{\theta}$ and $e(\theta) = e^R(\theta)$ for $\theta \geq \hat{\theta}$. Furthermore, because the solution satisfies $e(\theta) \leq e^R(\theta)$ for all θ and $e^R(\theta) < e^{FB}(\theta)$ for all θ, it follows that $e(\theta) < e^{FB}(\theta)$ for all θ.

In the event that there is no cut-off type $\hat{\theta} \in (\underline{\theta}, \bar{\theta})$ that solves (B7), there is complete pooling at some $\hat{e}$. To identify $\hat{e}$, the Pontryagin condition can again be integrated from $\underline{\theta}$ up to $\bar{\theta}$ and combined with the boundary conditions on $v(\underline{\theta})$ and $v(\bar{\theta})$ to yield:

$$\int_{\underline{\theta}}^{\bar{\theta}} S_e(\hat{e}, \theta)p(\theta)\, d\theta = \frac{\mu}{1 + \mu \dfrac{\delta}{1 - \delta}} \{c_e(\hat{e}, \bar{\theta})\}.$$

In this case $\hat{e} < e^{FB}(\bar{\theta})$, so $e(\theta) = \hat{e} < e^{FB}(\theta)$ for all θ.

Remark: The strong assumption that P is concave is used to guarantee that the separating schedule $e^R(\theta)$ defined by (B6) is decreasing. If the assumption is relaxed, a partial pooling optimum still involves pooling of the least efficient cost types but may also involve additional pooling regions. To derive the optimal contract, the boundary of the lower pooling interval, $\hat{\theta}$, is identified as before but above $\hat{\theta}$ the candidate schedule $e^R(\theta)$ must be "ironed" if it fails to satisfy monotonicity. It is still the case that all types choose less than efficient effort.

The assumption that $e^{FB}(\theta)$ is interior can also be relaxed. The only change is that even in a second-best environment $e(\theta)$ may equal $e^{FB}(\theta)$ for some cost types. Specifically, it may be that $e(\theta) = e^{FB}(\theta) = 0$ for very high costs or that $e(\theta) = e^{FB}(\theta) = \bar{e}$ for very low costs.

PROOF OF THEOREM 6:

The optimal stationary contract $e(\theta)$, $W(\theta, y) = w + b(\theta, y)$ solves:

$$\max_{e(\cdot), W(\cdot,\cdot)} \; s = \mathbb{E}_{\theta,y}[y - c \big| e(\theta)]$$

$$\text{subject to } \frac{d}{de}\{\mathbb{E}_y[W(\theta, y) - c(e, \theta) \big| e = e(\theta)]\} = 0 \text{ for all } \theta,$$

$$\frac{\delta}{1 - \delta}(s - \bar{s}) \geq \sup_{\theta,y} W(\theta, y) - \inf_{\theta,y} W(\theta, y).$$

In addition, the total compensation $W(\theta, y)$ must be structured so that neither party will want to renege on payments—this can be done as in Theorem 4. Here I have substituted the agent's first-order condition for the (*IC*) constraint. This is valid under the Mirrlees-Rogerson conditions. The result now follows from the fact that the optimization problem is linear in $W(\theta, y)$. I omit the details.

Appendix C: Subjective Performance Measures

I sketch the analysis with subjective performance measures. Details are contained in a secondary Appendix available from the author.

PROOF OF THEOREM 7:

Assume that an optimal full-review contract exists and generates surplus s^*. The argument is first to show that for any optimal full-review contract there is a termination contract that achieves the same expected surplus. The next step is to argue that the optimal termination contract takes a cut-off form.

1. The first observation is that, similar to Theorem 1, the set of payoffs that can be achieved with a self-enforcing full-review contract is equal to $\{(u, \pi) : u \geq \bar{u}, \pi \geq \bar{\pi} \text{ and } u + \pi \leq s^*\}$.
2. The second step is to argue constructively, along the lines of Theorem 2, that a termination contract can achieve the optimal surplus s^*. To do this, one considers an optimal full-review contract that in its initial period specifies payments w, $b(y)$, effort e, and equilibrium path continuation payoffs $u(y)$, $\pi(y)$, and continuation surplus $s(y) \equiv u(y) + \pi(y)$. A termination contract is then constructed with payments w^*, $b^*(y)$, effort e, continuation probabilities $\alpha^*(y)$, and (in the event of continuation) optimal continuation payoffs $u^* + \pi^* = s^*$. The idea is to choose the discretionary payments and continuation payoffs so as to exactly duplicate the incentive structure under the initial contract—in particular, so that $b^*(y) + \frac{\delta}{1-\delta}\alpha^*(y)(u^* - u) = b(y) + \frac{\delta}{1-\delta}u(y)$ and also $-b^*(y) + \frac{\delta}{1-\delta}\alpha^*(y)(\pi^* - \bar{\pi}) = -b(y) + \frac{\delta}{1-\delta}\pi(y)$. The fixed compensation w^* is chosen to ensure the per-period payoffs are u^* and π^*. Given this construction the termination contract will generate the same surplus as the original contract and be self-enforcing as a result of the original being self-enforcing.
3. The final step is to argue that the optimal continuation contract will take a one-step form. Similar to Theorem 6, this result follows from risk neutrality. To see that bad outcomes are followed by termination consider the following argument. To provide effort incentives, good outcomes must generate a high future payoff for the agent, and consequently a low future payoff for the principal. To provide reporting incentives, the principal must be indifferent between outcomes. This means that bad outcomes must result in a low future payoff for both parties. Optimization implies this low payoff should be as low as possible—hence termination.

REFERENCES

Abreu, Dilip. "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica*, November 1988, *56*(6), pp. 383–96.

Abreu, Dilip; Milgrom, Paul and Pearce, David. "Information and Timing in Repeated Partnerships." *Econometrica*, November 1991, *59*(6), pp. 1713–33.

Abreu, Dilip; Pearce, David and Stacchetti, Ennio. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring." *Econometrica*, September 1990, *58*(5), pp. 1041–63.

Altman Weil, Inc. *Compensation systems in private law firms*. Newtown Square, PA: Altman Weil Publications, 2000.

Athey, Susan and Bagwell, Kyle. "Optimal Collusion with Private Information." *RAND Journal of Economics*, Autumn 2001, *32*(3), pp. 428–65.

Atkeson, Andrew. "International Lending with Moral Hazard and Risk of Repudiation." *Econometrica*, July 1991, *59*(4), pp. 1069–89.

Baker, George; Gibbons, Robert and Murphy, Kevin J. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, November 1994, *109*(4), pp. 1125–56.

_____. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics*, February 2002, *117*(1), pp. 39–84.

Bernheim, B. Douglas and Whinston, Michael D. "Incomplete Contracts and Strategic Ambiguity." *American Economic Review*, September 1998, *88*(4), pp. 902–32.

Bull, Clive. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics*, February 1987, *102*(1), pp. 147–59.

Carver, Todd and Vondra, Albert A. "Alternative Dispute Resolution: Why It Doesn't Work and Why It Does." *Harvard Business Review*, May–June 1994, pp. 120–30.

Compte, Olivier. "Communication in Repeated Games with Imperfect Private Monitoring." *Econometrica*, May 1998, *66*(3), pp. 597–626.

Edlin, Aaron S. and Reichelstein, Stefan. "Hold-ups, Standard Breach Remedies, and Optimal Investment." *American Economic Review*, June 1996, *86*(3), pp. 478–501.

Endlich, Lisa. *Goldman Sachs: The culture of success*. New York: Alfred A. Knopf, 1999.

Fast, Norman and Berg, Norman. "The Lincoln Electric Company." HBS Case Series 9–376–028. Boston, MA: Harvard Business School Publishing, 1975.

Fudenberg, Drew; Holmstrom, Bengt and Milgrom, Paul. "Short-term Contracts and Long-term Agency Relationships." *Journal of Economic Theory*, June 1990, *51*(1), pp. 1–31.

Fudenberg, Drew; Levine, David I. and Maskin, Eric. "The Folk Theorem with Imperfect Public Information." *Econometrica*, September 1994, *62*(5), pp. 997–1039.

Green, Edward J. and Porter, Robert H. "Noncooperative Collusion under Imperfect Price Information." *Econometrica*, January 1984, *52*(1), pp. 87–100.

Grossman, Sanford and Hart, Oliver. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, August 1986, *94*(4), pp. 691–719.

Holmstrom, Bengt. "Moral Hazard and Observability." *Bell Journal of Economics*, Spring 1979, *10*(1), pp. 74–91.

Innes, Robert D. "Limited Liability and Incentive Contracting with *Ex Ante* Action Choices." *Journal of Economic Theory*, October 1990, *52*(1), pp. 47–65.

Kandori, Michihiro and Matsushima, Hitoshi. "Private Observation, Communication and Collusion." *Econometrica*, May 1998, *66*(3), pp. 627–52.

Klein, Benjamin and Leffler, Keith. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy*, August 1981, *89*(4), pp. 615–41.

Kreps, David M. "Corporate Culture and Economic Theory," in James E. Alt and Kenneth A. Shepsle, eds., *Perspectives on positive political economy*. Cambridge: Cambridge University Press, 1990, pp. 90–143.

Laffont, Jean-Jacques and Tirole, Jean. *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press, 1993.

Léonard, Daniel and van Long, Ngo. *Optimal control theory and static optimization in economics*. Cambridge: Cambridge University Press, 1992.

Levaggi, Rosella. "Optimal Procurement Contracts under a Binding Budget Constraint." *Public Choice*, October 1999, *101*(1–2), pp. 23–37.

Lowenstein, Roger. "Into Thin Air." *New York Times Magazine*, February 17, 2002, pp. 40–45.

Macaulay, Stewart. "Non-Contractual Relations in Business: A Preliminary Study." *American Sociological Review*, February 1963, *28*(1), pp. 55–67.

MacLeod, W. Bentley. "Optimal Contracting with Subjective Evaluation." *American Economic Review*, March 2003, *92*(1), pp. 216–40.

MacLeod, W. Bentley and Malcomson, James M. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica*, March 1989, *57*(2), pp. 447–80.

_____. "Motivation and Markets." *American Economic Review*, June 1998, *88*(3), pp. 388–411.

Macneil, Ian R. "Contracts: Adjustment of Long-Term Relationships under Classical, Neoclassical, and Relational Contract Law." *Northwestern University Law Review*, *72*, pp. 854–906.

Malcomson, James M. and Spinnewyn, Frans. "The Multiperiod Principal–Agent Problem." *Review of Economic Studies*, July 1988, *55*(3), pp. 391–407.

Pearce, David G. and Stacchetti, Ennio. "The In-

teraction of Implicit and Explicit Contracts in Repeated Agency." *Games and Economic Behavior*, April 1998, *23*(1), pp. 75–96.

Rayo, Luis. "Relational Team Incentives and Ownership." Working paper, University of Chicago Graduate School of Business, 2002.

Rogerson, William P. "The First-Order Approach to Principal-Agent Problems." *Econometrica*, November 1985, *53*(6), pp. 1357–67.

Sappington, David. "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory*, February 1983, *29*(1), pp. 1–23.

Schmidt, Klaus M. and Schnitzer, Monika. "The Interaction of Explicit and Implicit Contracts." *Economics Letters*, May 1995, *48*(2), pp. 193–99.

Shapiro, Carl and Stiglitz, Joseph E. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, June 1984, *74*(3), pp. 433–44.

Stewart, James B. "Taking the Dare." *The New Yorker*, July 26, 1993, pp. 34–39.

Thomas, Jonathan and Worrall, Tim. "Self-Enforcing Wage Contracts." *Review of Economic Studies*, October 1988, *55*(4), pp. 541–53.

Williamson, Oliver E. *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press, 1985.