TECHNICAL CHANGE, MORAL HAZARD, AND THE DECENTRALIZATION PENALTY

Thomas Marschak[†] and Dong Wei[††]

July 23, 2020

In a moral-hazard setting, we consider two modes of organizing a firm — a Centralized mode and a Decentralized mode. In the Centralized mode, perfect monitoring ensures that the welfare-maximizing effort is chosen. In the Decentralized mode, monitoring is imperfect and a profit-driven Principal induces her favorite effort by appropriately rewarding the Agent. The loss in welfare due to decentralization is called the *Decentralization Penalty*. For certain common contract types, we study the behavior of the Decentralization Penalty in response to changes in the production technology. We find that as production technology improves, the Decentralization Penalty oscillates. The Penalty rises in all intervals where the Penalty changes continuously. After such an interval, there may be a sudden change in the Penalty, and, under reasonable assumptions on costs and expected revenues, the sudden change is a drop. While an improvement in monitoring technology always strengthens the case for the Centralized mode, advances in production technology may do the opposite.

† Haas School of Business, University of California, Berkeley.
†† Department of Economics, University of California, Berkeley.

1. Introduction.

How do the striking technical advances that we regularly observe affect the way firms are organized? The question is complex and hard to frame in a precise way. Here we seek insights from a simple model, where there are two possible modes of organizing.

One organizational mode is "Centralization," where perfect monitoring and policing ensure that the effort each subordinate makes is what the firm wants. That may be very expensive, especially if the subordinate has to be compensated to make up for the unpleasantness of the monitored working environment. Dramatic advances in monitoring technology have made it less expensive.[1] The other mode that we study is "Decentralization," where a manager plays the role of Principal, and each subordinate is an Agent who freely chooses his effort. The Agent bears the cost of the chosen effort, but is rewarded by the Principal. Advances in *production technology* lower the cost of a given effort. We consider a standard moral-hazard setting. An Agent's reward depends on the revenue which her effort yields. Revenue is uncertain, but the probability distribution of revenue is common knowledge. We study the case of a single Agent. In our setting there is a finite collection of possible efforts. An effort of level e costs tC_e, where C_e is strictly increasing in e. When production technology improves, t drops.

We compare the two modes of organizing with regard to welfare, which we may also call *surplus*. That can, for example, be the comparison made by a government agency which regulates the firm. At the technology level t and effort e, welfare (surplus) is

$$[\text{expected revenue at effort } e] \ - tC_e.$$

In the Centralized mode the firm chooses the (largest) effort that maximizes surplus. We let $\gamma(t)$ denote that effort. In the Decentralized mode, we assume that both Principal and Agent are risk-neutral and that a limited-liability requirement is satisfied: the Agent's reward is never negative. If the Principal wants to induce a particular effort, she uses an appropriate *contract*, i.e., a vector of nonnegative rewards, one for each of the possible revenues. The Principal chooses a *favorite* effort, denoted $\delta(t)$, which is the largest maximizer of the expected profit

$$[\text{expected revenue at } e] - [\text{the smallest expected reward that induces the Agent to choose } e].$$

Our welfare-oriented judge of the two modes is interested in the *Decentralization Penalty*, denoted $D(t)$. That is the difference between maximal surplus and the surplus in the Decentralized mode. Thus

$$D(t) = \left[\text{expected revenue at } \gamma(t) \ - tC_{\gamma(t)}\right] \ - \ \left[\text{expected revenue at } \delta(t) - tC_{\delta(t)}\right].$$

[1] A recent survey of workplace monitoring and surveillance is Mateescu and Nguyen(2019). Legal aspects are surveyed in Tippet *et al*, (2017) and Vagle (2020). Zuboff (2019) studies the limits on workplace surveillance imposed by political and legal institutions. Ball (2010) surveys research on workplace surveillance by social psychologists. Sophisticated monitoring and surveillance are reported to occur in the "fulfillment centers" operated by Amazon. See, for example, "I worked at an Amazon Fulfillment Center; They Treat Workers Like Robots", *TIME*, July 18, 2019.

Here is our central question: *does D rise or fall when production technology improves (t drops)?* If it falls, then the case for Decentralization — when we are interested in welfare — becomes stronger. If it rises, the case for Decentralization weakens, and if it rises sufficiently, then Centralization is preferred, despite the cost of perfect monitoring/policing. One can ask a related question: what happens if monitoring technology improves? Let K be the social cost of perfect monitoring — the value of the resources required to guarantee that the effort is indeed welfare-maximizing. Our welfare-oriented judge prefers the Decentralized mode if $D(t) < K$. When monitoring technology improves (K drops), the superiority of the Decentralized mode — for a fixed t — decreases. If monitoring technology improves sufficiently, the Centralized mode becomes superior.

We obtain sharp results about the behavior of the Decentralization Penalty when we impose appropriate restrictions on the Principal's contracts. We consider two contract types: *bonus* contracts and *fixed-share* contracts, both of which are widely studied in the literature and used in practice. In a bonus contract, the Agent receives a positive reward only when the highest possible revenue is realized. In a fixed-share contract, the Agent always receives a fixed share of the realized revenue.

We find that under either type of contract, the Decentralization Penalty oscillates with production technology in a particular way. When the cost of production is sufficiently high, the Penalty is zero because it is both welfare- and profit-maximizing to induce the lowest effort. As production technology advances (the production cost parameter t drops), the Penalty starts to be positive and to increase. We will see that, under both bonus and fixed-share contracts, the Principal never *squanders*: he never induces an effort higher than the welfare-maximizing level (i.e., $\delta(t) \leq \gamma(t), \forall t$). As a result, the Penalty becomes positive when the effort induced by a profit-driven Principal is too low compared to the welfare-maximizing level; when this happens, although some production cost is saved by inducing a lower effort, that saving in cost is outweighed by the associated loss in expected revenue. Consequently, on any interval where the Penalty is positive while $\delta(t)$ and $\gamma(t)$ stay constant,[2] the Penalty smoothly *increases* as production technology improves, because the saving in cost caused by a lower effort gradually decreases while the loss in expected revenue remains. Once in a while, however, there will be a sudden *drop* in the Decentralization Penalty, because at that point the profit-maximizing effort jumps to a larger member of the finite set of efforts, so it gets closer to the welfare-maximizing effort. This continuous-rise-sudden-drop cycle repeats itself until the production cost is so low that the welfare- and profit-maximizing efforts both reach the highest level, so the Penalty is again zero.

To obtain these results, we make a natural assumption about costs and expected revenues: when effort increases, surplus rises until it reaches a peak (where it may stay the same for a while) and then decreases. If we drop that assumption (which we call *quasi-concavity*), then it remains true that the Penalty never smoothly decreases when t drops. But it may now happen that there are sudden *rises* in the Penalty as well as sudden drops, so the continuous-rise-sudden-drop cycle becomes a more general continuous-rise-sudden-change cycle.

The remainder of the paper is organized as follows. Below we review the literature. Section 2 presents the model. Section 3 establishes some basic properties of optimal efforts and the Decentralization Penalty. Sections 4 and 5 consider bonus contracts and fixed-share contracts, respectively,

[2] Note that δ and γ are step functions because there is a finite number of efforts.

and characterize the way the Penalty behaves as technology improves. Section 6 considers some variations of our results under alternative assumptions. Section 7 considers some economic applications. Section 8 concludes. An Appendix provides proofs.

Related literature. The paper closest to ours is Balmaceda *et al* (2016). That paper concerns the same moral-hazard model that we consider here, but it studies a fraction which computer scientists call "the price of anarchy". The numerator is the highest possible surplus (welfare) and the denominator is the lowest surplus that can occur when the Principal induces his favorite effort.[3] Their main aim is to find a useful upper bound to this ratio. Under the assumptions made, the number of efforts turns out to be an upper bound. The effect of effort-cost reduction on the ratio is not examined. Balmaceda *et al* also show that the Principal loses nothing by confining attention to bonus contracts if the revenue distributions have the monotone-likelihood-ratio proprty (a standard assumption which we also make), and in addition a condition called "Increasing Marginal Cost of Probability" (IMCP) is satisfied.[4] In our setting, it is natural to focus on the surplus gap, rather than the ratio considered in Balmaceda *et al.* Perfect monitoring (Centralization) would eliminate the gap, but it can be costly. If such costs exceed the surplus gap, then Decentralization is the preferred mode. There are no simple analogs of those statements if we define the Penalty to be a fraction with maximal surplus in the numerator and Decentralized surplus in the denominator.

Schmitz (2005) considers an Agent who can choose any effort in the interval $[0, 1]$. The resulting revenue (in our terminology) can be High or Low, with probabilities that are common knowledge. The Agent is rewarded by a Principal once revenue is realized. But if the Principal spends money on a surveillance device, then the Agent's effort choice becomes observable and the Agent is rewarded (or punished) when the choice is made. It is shown that if the surveillance cost is sufficiently high, then welfare increases if the device is banned. Another model related to ours is studied in Acemoglu and Wolitzky (2011). The novelty of the model is that the Agent is punished if he refuses to participate and leaves. So there is a family of organizational modes rather than our two modes; each mode corresponds to a possible punishment level.[5] It is shown that increasing the exit punishment (or lowering the cost of a given punishment level) reduces welfare.

Turning to the general Principal/Agent problem, we find many papers, starting with the earliest ones, where the Agent's effort may have a cost. The Agent has a utility function on her actions

[3] In many price-of-anarchy papers the fraction studied has the socially best outcome as the denominator (unlike the fraction in Balmaceda *et al*). The typical paper concerns a game. The fraction's denominator might be the largest possible payoff sum, which is attainable if the players cooperate. The numerator is the payoff sum in the "socially worst" equilibrium of the game. A variety of social situations have been studied from this point of view. One of them concerns optimal versus "selfish" routing in transportation (Roughgarden (2005)). Others are found in Nissan, Roughgarden, Tardos, and Vazirani (eds.) (2007). Many of these studies develop bounds on the price of anarchy. Several of them (e.g., Babaioff, Feldman, and Nissan (2009)) consider a Principal/Agent setting. Moulin (2008) considers various cost-sharing games (where each player demands a quantity of a good and is charged a share of the cost of meeting all players' demands) and studies the price of anarchy in each of them.

[4] IMCP concerns a fraction. The fraction is a function of effort level. Its numerator is the extra cost of going one level higher and its denominator is the resulting increase in the probability of the highest revenue. IMCP says that the fraction is (weakly)increasing in the level. In Section 4.1 we discuss IMCP in more detail.

[5] One interpretation of the model is slavery, and that is indeed the focus of the paper. There are just two revenues and the set of possible efforts is $[0, 1]$.

and rewards. Agent utility for the action a and the reward y takes the form $V(y) - g(a)$. Among the early papers where this occurs are Holmstrom (1979), (1982) and Grossman and Hart (1983). The action a might be effort and $g(a)$ could be its cost. Welfare loss also appears very early in the literature. Ross (1973), for example, finds conditions under which the solution to the Principal's problem maximizes welfare (as measured by the sum of Agent's utility and Principal's utility) and notes that these conditions are very strong. But Principal/Agent papers whose main concern is the relation between effort cost and welfare loss appear to be scarce.

2. The Model

There is one risk-neutral worker, whom we refer to as the Agent. There are $E \geq 2$ possible efforts. They are denoted $1, 2, \ldots, e, \ldots, E$, where each is higher than its predecessor. The cost of effort e is tC_e, where $C_e > 0$ and $t > 0$. C_e is strictly increasing in e, and $t > 0$ is a technology parameter which drops when technology improves. There are $S \geq 2$ possible revenues, denoted $R_1, R_2, \ldots, R_S$, where $0 \leq R_1 < R_2 < \cdots < R_S$. The probability distribution of revenue depends on the effort chosen. When effort e is chosen, the probability that revenue will turn out to be R_s is p^e_s. We let p^e denote the vector $(p^e_1, \ldots, p^e_S)$. So our problem is defined by the triple $(\{C_e\}_{e=1,\ldots,E}\ ,\ \{R_s\}_{s=1,\ldots,S}\ ,\ \{p^e\}_{e=1,\ldots,E})$.

For effort e, we let $\overline{R}^e$ denote the average revenue (i.e., $\overline{R}^e = \sum_{s=1}^{S}\ p^e_s R_s$). Then surplus is $\overline{R}^e - tC_e$.

As is common in the moral-hazard literature, we shall assume that the revenue distributions have the *Monotone-Likelihood-Ratio (MLR)* property.

Definition 1

The probabilities $\{(p^e_1, \ldots, p^e_S)\}_{e=1,\ldots,E}$ have the Monotone Likelihood Ratio (MLR) property if

$$\frac{p^e_{s^*}}{p^e_s} > \frac{p^f_{s^*}}{p^f_s} \text{ whenever } e > f, s^* > s.$$

Informally: when effort increases, so does the ratio of the probability that we will see a given revenue to the probability that we will see a lower one. MLR implies that if $e > f$ the cumulative distribution function of revenue for effort e first-order stochastically dominates the cumulative distribution function for effort f. Consequently

- Expected revenue strictly increases when effort increases;
- The probability p^e_S of the highest revenue R_S strictly increases when effort increases.

We shall consider a second assumption, which we call *Quasi-concavity*

Definition 2

The triple $(\{C_e\}_{e=1,\ldots,E}\ ,\ \{R_s\}_{s=1,\ldots,S}\ ,\ \{p^e\}_{e=1,\ldots,E})$. has the *Quasi-concavity* property if surplus is a unimodal function of effort. There may be more than one surplus-maximizing effort. Surplus drops whenever effort falls further below the smallest surplus-maximizing effort. [6]

In the *Centralized mode*, monitoring guarantees that the chosen effort maximizes surplus.[7] In the *Decentralized mode*, efforts are freely chosen by a risk-neutral Agent, who is rewarded by a risk-neutral Principal but bears the cost of the chosen effort. The Principal offers the Agent a *contract* $w = (w_1, \ldots, w_S)$. The *wage* $w_s \geq 0$ is paid to the Agent if revenue turns out to be R_s. We let $\overline{w}^e$ denote the average wage received if the Agent indeed chooses the effort e. Thus $\overline{w}^e = \sum_{s=1}^{S} p_s^e w_s$. If the Agent willingly chooses e, then w must satisfy two conditions:

- $\overline{w}^e - tC_e \geq 0$. This is the Individual Rationality (IR) requirement. It says that the Agent's expected net gain when he chooses effort e is at least zero, which is his net gain from the best outside option.

- $\overline{w}^e - tC_e \geq \overline{w}^f - tC_f$ for all $f \neq e$. This is the Individual Compatibility (IC) requirement. It says that the Agent's gain from choosing effort e is not less than the net gain from any other effort.

If a contract w satisfies the IR and IC conditions for the effort e, we shall say that it *induces* e. The Principal, like the Agent, is risk-neutral. Accordingly we informally say that a contract w which induces e *costs* the Principal $\overline{w}^e$. There may be many contracts that induce e. Among them the Principal seeks a contract that is cheapest. To find it, he solves a linear-programming problem which we shall call *the optimally-induce-e problem*:

> Find a vector $w = (w_1, \ldots, w_S)$ of nonnegative wages which minimizes $\overline{w}^e$ subject to the IR and IC constraints.

We have no interest in efforts that cannot be optimally induced. So henceforth *we restrict attention — without further comment — to triples $\left(\{C_e\}_{e=1,\ldots,E}\ ,\ \{R_s\}_{s=1,\ldots,S}\ ,\ \{p^e\}_{e=1,\ldots,E}\right)$ which have the property that for every effort e, the optimally-induce-effort-e problem has a solution.*

If w solves the optimally-induce-effort-e problem for a given t, we let the symbol $A_e(t)$ denote the average wage $\overline{w}^e$. So $A_e(t)$ is the lowest cost of inducing e. The Principal (weakly) prefers the effort e to the effort f if her net gain is not lower for e, i.e., $\overline{R}^e - A_e(t) \geq \overline{R}^f - A_f(t)$. We now formally define the two efforts which we introduced in the previous section: the welfare-maximizing effort and the Principal-favorite effort.

Definition 3

[6]Quasi-concavity is usually defined for functions whose domain is a convex space. In our model the effort space is finite. Formally, we can interpret our assumption in the following way: there exists a quasi-concave extension of the surplus function to the interval $[1, E]$. Quasi-concavity of surplus is implied, for example, by weak concavity of expected revenue and weak convexity of cost with respect to effort.

[7]If there is a Principal, she is able to induce the Agent to choose the welfare-maximizing effort even though his compensation when he does so equals the cost he incurs.

Given t, the **welfare-maximizing effort**, denoted $\gamma(t)$, is the largest element in

$$\text{argmax}_{e\in\{1,\ldots,E\}} \left[\overline{R}^e - tC_e\right].$$

The **Principal-favorite effort**, denoted $\delta(t)$, is the largest element in

$$\text{argmax}_{e\in\{1,\ldots,E\}} \left[\overline{R}^e - A_e(t)\right].$$

Note that the functions $\gamma(\cdot), \delta(\cdot)$ are (integer-valued) step functions. Each takes at most E values. Note also that ties, if any, are broken upward. [8]

The *Decentralization Penalty*, denoted $D(t)$, is the difference between the maximal surplus and the surplus in the Decentralized mode under the Principal-favorite effort. So

$$D(t) = [\overline{R}^{\gamma(t)} - tC_{\gamma(t)}] - [\overline{R}^{\delta(t)} - tC_{\delta(t)}] = [\overline{R}^{\gamma(t)} - \overline{R}^{\delta(t)}] + t\cdot\left(C_{\delta(t)} - C_{\gamma(t)}\right). \tag{1}$$

Note that the Penalty is never negative, and is equal to zero if $\gamma(t) = \delta(t)$. The behavior of the function $D(\cdot)$ is our main concern.

3. Basic Properties of the Decentralization Penalty

In this section we establish two basic properties of optimal efforts and the Decentralization Penalty. Those properties will be useful in the subsequent analysis.

3.1 Monotonicity of Welfare-Maximizing and Principal-Favorite Efforts

When production is less costly, it becomes easier to induce the worker to choose a higher effort. Consequently, the welfare-maximizing and principal-favorite efforts rise with technological improvement. That is established in the following Lemma.

Lemma 1

If $t^{**} < t^*$, then $\gamma(t^{**}) \geq \gamma(t^*)$ and $\delta(t^{**}) \geq \delta(t^*)$.

The inequality concerning the welfare-maximizing effort $\gamma(\cdot)$ is easy to prove. Consider efforts e, f, where $e > f$. Note that if, at t^*, welfare at effort e is weakly higher than welfare under f, then

$$\overline{R}^e - \overline{R}^f \geq t^*\cdot(C_e - C_f).$$

Since $C_e > C_f$, it must be the case that at $t^{**} < t^*$ we have

$$\overline{R}^e - \overline{R}^f > t^{**}\cdot(C_e - C_f).$$

So the welfare ranking of the two efforts cannot be reversed when t drops. Hence an effort that maximizes welfare at t^{**} cannot be lower than an effort that maximizes welfare at t^*.

[8] In our main analysis, we find $\gamma(t) \geq \delta(t)$. So if we assume quasi-concavity of surplus, then the upward tie-breaking rule for δ coincides with another tie-breaking rule, where the Principal always breaks ties in favor of social welfare.

The inequality that concerns the Principal-favorite effort $\delta(\cdot)$ is of interest in itself. It says that a technical advance never leads the Principal to induce less effort from the Agent. Its proof in the Appendix exploits the Strong Duality theorem of linear programming.

3.2 Squandering and the Decentralization Penalty

We say the Principal *squanders at t* if $\delta(t) > \gamma(t)$. Recall the definition of Decentralization Penalty in (1). Since C_e is strictly increasing in e, the following must hold:

Lemma 2

Suppose that the step functions $\gamma(\cdot)$ and $\delta(\cdot)$ are constant throughout an interval, i.e., there exist γ', δ' such that $\gamma(t) = \gamma', \delta(t) = \delta'$ at all t in the interval. Then in that interval D is strictly decreasing if $\delta' < \gamma'$, strictly increasing if $\delta' > \gamma'$ (squandering), and zero if $\delta' = \gamma'$.

Any interval on which the step functions $\gamma(\cdot)$ and $\delta(\cdot)$ are not equal has a subinterval where each is constant. So Lemma 2 will be an important tool in characterizing the behavior of the Penalty. If we graph the Penalty, with t on the horizontal axis, and we move from right to left (technology improves), then if there is no squandering we will not find an interval where the Penalty continuously falls. But there may be critical values of t where the Penalty suddenly changes. We shall see that bonus contracts and fixed-share contracts, studied in the next two sections, have the "no-squandering" property.

4. Bonus Contracts

It is common practice to pay a worker a fixed amount and to add a bonus that depends on the observed result of the worker's unobserved effort. The term "bonus" appears often in the moral-hazard literature.[9] In our setting, we shall call a contract $w = (w_1, \ldots, w_S)$ a *bonus contract* if it has the form $(0, \ldots, 0, z)$, where $z > 0$. The Agent receives a fixed amount (normalized at zero) regardless of revenue and is rewarded with a positive bonus when revenue turns out to be R_S, the highest possible revenue. In this section, we characterize the behavior of the Decentralization Penalty under bonus contracts.

Before proceeding we ask whether the restriction to bonus contracts is without loss as far as profit maximization is concerned. As previously noted, Balmaceda *et al* introduce a condition on the effort costs and the revenue distribution which they call Increasing Marginal Cost of Probability (IMCP). To understand this condition, recall that MLR implies that the probability of the highest revenue R_S is strictly increasing in effort. IMCP concerns the ratio of the marginal cost of exerting a higher effort to the resulting increase in the probability of R_S. The assumption requires that ratio to be weakly increasing in effort.[10] That happens, for example, if the marginal cost of effort is weakly increasing and there are decreasing marginal returns to effort in the following sense: the improvement in the

[9]For example, Herweg *et al* (2010) use the word "bonus" to describe a contract where the Agent receives a lump sum if a certain level of performance is reached. In papers where there are just two possible outcomes (e.g., Halac *et al* (2016), Fong and Li (2016)), the reward for a "success" may be labeled a "bonus". However, there appears to be no standard definition of the term.

[10]Formally, let $v_1 \equiv \frac{C_1}{p_1^S}$ and $v_e \equiv \frac{C_e - C_{e-1}}{p_S^e - p_S^{e-1}}$ for $2 \leq e \leq E$. IMCP is satisfied if $v_1 \leq v_2 \leq \cdots \leq v_E$.

probability of the highest revenue decreases with effort.[11] Balmaceda *et al* show (in their Propositon 1) that if IMCP is satisfied, then the Principal-favorite contract has the bonus form, so we lose no generality when we confine attention to bonus contracts. [12]

4.1 Inducing an Effort with a Bonus Contract

We shall say that a contract *bonus-induces* an effort e if it has the bonus form and satisfies the IR and IC conditions. We shall say that a contract optimally bonus-induces e if it bonus-induces e and does so at least as cheaply as any other contract which bonus-induces e. To find a contract which optimally bonus-induces a given effort, say e, the Principal solves a linear programming problem which is simpler than the optimally-induce-e problem when contracts are unrestricted. The new problem is:

Find nonnegative z which minimizes $p_S^e z$ subject to

$$p_S^e z \geq tC_e \qquad \text{(IR)}$$

$$p_S^e z - tC_e \geq p_S^f z - tC_f \text{ for all } f \neq e \qquad \text{(IC)}$$

Henceforth we confine attention to triples $\left(\{C_e\}_{e=1,\ldots,E}, \{R_s\}_{s=1,\ldots,S}, \{p^e\}_{e=1,\ldots,E}\right)$ which have the property that the linear programming problem has a solution, so each of the m possible efforts can be optimally bonus-induced. Let $z^e(t)$ be the solution to the problem and let $A_e^b(t)$ denote $p_S^e z^e(t)$. Thus it costs the Principal $A_e^b(t)$ to optimally bonus-induce the effort e.

Does it cost the Principal more to induce a higher effort? The answer is in the affirmative for bonus contracts.[13] To see this, suppose that $f > e$. To bonus-induce effort f, the solution $z^f(t)$ to the linear programming problem must satisfy

$$z^f(t) \geq \frac{C_f - C_e}{p_S^f - p_S^e}.$$

On the other hand, to bonus-induce effort e, the solution $z^e(t)$ must satisfy

$$z^e(t) \leq \frac{C_f - C_e}{p_S^f - p_S^e}.$$

Thus $z^f(t) > z^e(t)$ and $A_f^b(t) > A_e^b(t)$ (since $p_S^f > p_S^e$, by MLR).

[11] Balmaceda *et al* also point out that IMCP is weaker than the well-known "convexity of the distribution function condition" which is often used to validate the first-order approach in the continuous-effort case.

[12] In our setting the Balmaceda *et al* result means that when he induces the effort e, the Principal loses nothing by confining attention to contracts which have the form $(0, \ldots, 0, tv_e)$. We shall see in section 4.3 that if there are just two efforts, then IMCP is not only a sufficient condition for the Principal to lose nothing by confining attention to bonus contracts; it is necessary as well.

[13] As we shall see in section 6.1, that is not necessarily true when contracts are unrestricted.

For brevity, we slightly abuse the symbol δ. We now use $\delta(t)$ to denote the *Principal-favorite effort at t under bonus contracts*. That is defined as the largest member of the set

$$\mathrm{argmax}_{e\in\{1,\ldots,E\}}\left[\overline{R}^e - A^b_e(t)\right].$$

So, as in Definition 3, ties are broken upward. As before, $\gamma(t)$ denotes the largest welfare-maximizing effort, and the Decentralization Penalty at t is

$$D(t) = [\overline{R}^{\gamma(t)} - \overline{R}^{\delta(t)}] + t\cdot(C_{\delta(t)} - C_{\gamma(t)}).$$

We now have a lemma which is an analog of Lemma 1. It concerns the Principal-favorite effort when bonus contracts must be used.

Lemma 2

If bonus contracts must be used, then $t^{**} < t^*$ implies $\delta(t^{**}) \geq \delta(t^*)$.

The proof again uses strong duality.

4.2 The Behavior of the Decentralization Penalty when bonus contracts are used

We now turn to the behavior of the Decentralization Penalty for bonus contracts. We start with a lemma, which will be an important tool.

Lemma 3

If bonus contracts must be used, then $\delta(t) \leq \gamma(t)$ for all $t > 0$. (The Principal never squanders).

If we now make the MLR and Quasi-concavity assumptions, then the following proposition fully characterizes the behavior of the Decentralization Penalty when bonus contracts are used.

Proposition 1.

Suppose the Monotone Likelihood Ratio property and the Quasi-concavity property hold. Suppose the Principal is required to use a bonus contract. The Penalty is zero for sufficiently small and sufficiently large t. Either the Penalty is zero at all t, or else there are intervals of finite length in each of which $D(t)$ begins with an upward jump to a positive number. After the jump, the Penalty decreases linearly until the end of the interval. Thus on each of these intervals the Penalty increases linearly as production technology advances (as t drops).

Formally, either $D(t) = 0$ at all t, or else there exist numbers $t_1,\ldots,t_q,\ldots,t_Q$ such that $Q > 1$; $0 < t_1 < \cdots < t_Q$; and the following statements hold:

- $D(t) = 0$ for $t \leq t_1$ and for $t \geq t_Q$.
- For $q \in \{1,\ldots,Q-1\}$ and $t \in (t_q, t_{q+1}]$, there exist G_q, u_q such that

$$D(t) = \frac{G_q - u_q}{t_{q+1} - t_q}\cdot(t_{q+1} - t) + u_q$$

  and either $G_q = u_q = 0$ or $G_q > u_q \geq 0$. Moreover, $G_{q+1} \geq u_q$.

The proposition says that if t is sufficiently high then the lowest possible effort is both welfare-maximizing and Principal-favorite, so the Penalty is zero. As production technology improves (t falls), the Decentralization Penalty varies in a continuous-rise-sudden-drop cycle until the production cost is low enough that the highest possible effort is both welfare-maximizing and Principal-favorite, so the Penalty is again zero. Intuitively, by Lemma 4, the Penalty is positive if and only if the Principal-favorite effort is strictly lower than the welfare-maximizing effort. When this happens, the loss in expected revenue due to a lower effort outweighs the saving in production cost. As a result, on any interval where the Penalty is positive while $\delta(t)$ and $\gamma(t)$ stay constant, the Penalty increases as production technology improves (t drops), because the saving in cost from inducing a lower effort ($t \cdot [C_{\gamma(t)} - C_{\delta(t)}]$) gradually decreases while the loss in expected revenue ($\overline{R}^{\gamma(t)} - \overline{R}^{\delta(t)}$) remains. Once in a while, however, there will be a sudden drop in the Decentralization Penalty, because at those points the Principal-favorite effort moves to a higher level and thus gets closer to the welfare-maximizing effort. Such a continuous-rise-sudden-drop cycle continues until production cost is so low that the welfare-maximizing and Principal-favorite efforts are both the highest possible effort.

We show in the proof that if we have quasi-concavity then as production technology improves, the only possible jump in the Penalty is indeed a downward jump. Inspecting the proof, we will see that when quasi-concavity is dropped (but MLR is retained) an upward jump cannot be excluded. Then the "continuous-rise-sudden-drop" cycle can no longer be claimed. Instead we may claim a more complex "continuous-rise-sudden-change" cycle.

We illustrate the typical shape of $D(t)$ with the following example.

Example 1

There are three efforts and three revenues ($E = S = 3$). The revenue distributions for the three efforts are as follows:

	R_1	R_2	R_3
effort 1	$p_1^1 = \frac{3}{10}$	$p_2^1 = \frac{2}{10}$	$p_3^1 = \frac{5}{10}$
effort 2	$p_1^2 = \frac{2}{10}$	$p_2^2 = \frac{2}{10}$	$p_3^2 = \frac{6}{10}$
effort 3	$p_1^3 = \frac{1}{10}$	$p_2^3 = \frac{1}{10}$	$p_3^3 = \frac{8}{10}$

Let $(C_1, C_2, C_3) = (2, 3, 9)$ and let $(R_1, R_2, R_3) = (1, 3, 10)$.

Note that both MLR and IMCP are satisfied.[14] The Penalty is graphed in Figure 1.[15]

[**FIGURE 1 HERE**]

In the figure we label the values of γ and δ in five intervals. When we move from right to left, these values do not decrease. We find two points of t where the Penalty suddenly drops to zero. Each drop is indicated by broken vertical lines. Each drop is preceded by an interval where the Principal does not squander and the Penalty linearly rises as t decreases. Thus, proceeding from *right to left*,

[14] We have $v_1 = C_1/p_3^1 = 4$, $v_2 = (C_2 - C_1)/(p_3^2 - p_3^1) = 10$, and $v_3 = (C_3 - C_2)/(p_3^3 - p_3^2) = 30$. So IMPC is satisfied

[15] The Decentralization Penalty is calculated for all t in $\left\{\frac{1}{100}, \frac{2}{100}, \cdots, \frac{99}{100}, 1\right\}$.

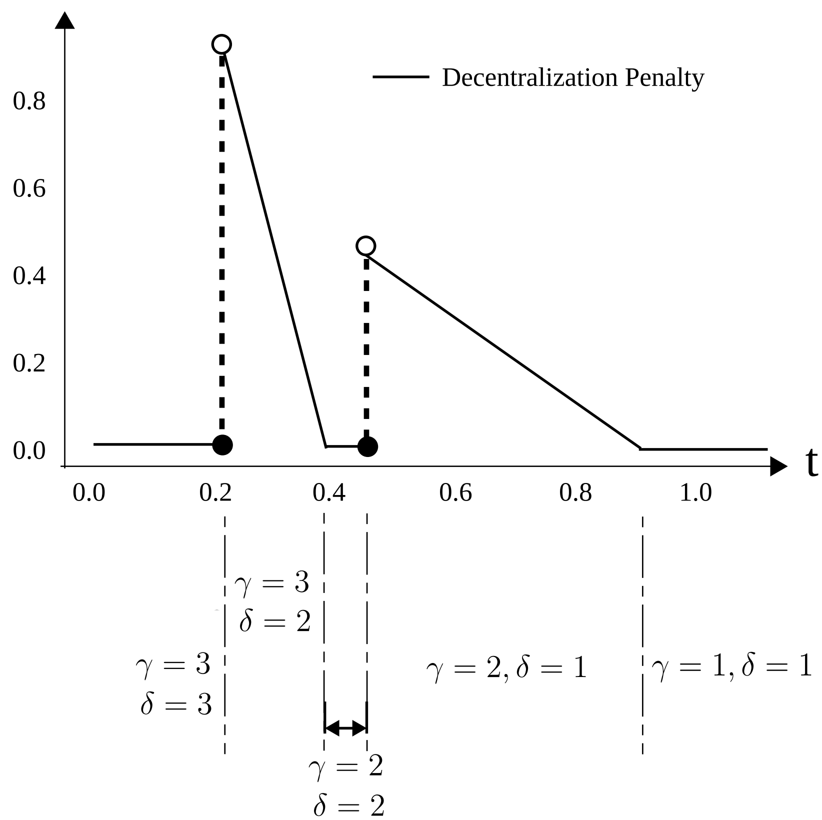

Figure 1 (Example 1)

Figure 1 illustrates the continuous-rise-sudden-drop cycle of the Decentralization Penalty for bonus contracts when we assume both MLR and Quasi-concavity.

4.3 Bonus contracts in the two-effort case.

As we have noted, Balmaceda *et al* show that if we make the IMPC assumption (as well as MLR), then the Principal loses nothing when she confines attention to bonus contracts. The IMCP requirement is strong. In the case of two efforts, however, we can claim that IMCP is both necessary and sufficient for the Principal to lose nothing by confining attention to bonus contracts.

To make that informal claim more precise, let us confine attention to cases where each of the efforts can be optimally induced (with no restriction on contract type) and can also be optimally bonus-induced. Then the Principal loses nothing by confining attention to bonus contracts if for every pair (e,t), with $e \in \{1,2\}$ we have $A_e^b(t) \leq A_e(t)$. The cost of a contract that bonus-optimally induces e never exceeds the cost of an unrestricted contract that optimally induces e.

We now claim that

(2) If each of the two efforts can be optimally induced and can also be bonus-optimally induced, then IMPC holds if and only if a contract that bonus-optimally induces e never costs more than an unrestricted contract that optimally induces e.

The argument has two parts.

Inducing effort 1:

A simple argument shows that the effort x_1 can be optimally induced with the non-bonus contract $(tC_1, tC_1, \ldots, tC_1)$, where the Principal always pays the Agent tC_1 and so the Agent's net gain is zero. The only candidate for a bonus contract which gives the Principal the same net gain (for any revenues) as $(tC_1, tC_1, \ldots, tC_1)$ is

$$w^* = \left(0, 0, \ldots, 0, \frac{tC_1}{p_S^1}\right).$$

That has the same cost for the Principal as $(tC_1, tC_1, \ldots, tC_1)$. The IR requirement for inducing x_1 is satisfied and binds. But the IC requirement for w^* is $0 \geq p_S^2 \cdot \dfrac{tC_1}{p_S^1} - tC_2$ or

$$C_2 p_S^1 \geq C_1 p_S^2. \tag{3}$$

We can find C_1, C_2 and probabilities having the MLR property for which that inequality is violated.[16]

On the other hand, the IMCP condition $v_1 \leq v_2$ is equivalent to[17] (3). So if IMCP is violated, then effort 1 cannot be optimally bonus-induced, but if IMCP is satisfied, then w^* optimally bonus-induces effort 1.

Inducing effort 2:

[16] Let $C_1 = 1$ and let $C_2 = p_S^2/p_S^1 + \epsilon$, where $\epsilon > 0$.

[17] Recall that $v_1 = C_1/p_S^1, v_2 = (C_2 - C_1)/(p_S^2 - p_S^1)$.

Balmaceda *et al* show that if IMPC holds, then effort 2 is optimally induced by the bonus contract

$$\tilde{w} = (0, 0, \ldots, 0, tv_2) = \left(0, 0, \ldots, 0, t \cdot \frac{C_2 - C_1}{p_S^2 - p_S^1}\right).$$

The IR condition for effort 2 is equivalent to (2). The IC condition

$$p_S^1 \cdot \frac{C_2 - C_1}{p_S^2 - p_S^1} - p_S^2 \cdot \frac{C_2 - C_1}{p_S^2 - p_S^1} \geq C_2 - C_1$$

holds as an identity. To show optimality of $\tilde{w}$, we have to consider any contract — say $w' = (w_1', \ldots, w_S')$ — which also meets the induce-effort-2 IR and IC conditions, and we have to show that w' is not cheaper than $\tilde{w}$, i.e., $\overline{w'}^2 \geq \overline{\tilde{w}}^2$. Balmaceda *et al* provide an argument that uses MLR as well as IMCP and shows that we can construct a new contract which also meets IR and IC, does not cost more than w', and has a component that equals zero. Applying this procedure to each of the first $S-1$ components, one at a time, we end up with a bonus contract, whose average wage cannot be higher than $\overline{w'}^2$.[18]

That completes the proof of statement (2).

5. Fixed-share Contracts

We now turn to another type of simple contract. A contract w has the fixed-share property if

$$w = (rR_1, rR_2, \ldots, rR_S),$$

where $0 \leq r \leq 1$. At a given t, the Principal chooses r and the Agent responds by choosing the effort he finds best. The Principal chooses r so as to maximize her expected payoff.[19]

[18] The argument is much simpler for the two-revenue case ($S = 2$). Any contract $w = (w_1, w_2)$ which induces x_2 satisfies the IC condition

$$w_1 \cdot (p_1^2 - p_1^1) + w_2 \cdot (p_2^2 - p_2^1) \geq t \cdot (C_2 - C_1),$$

which we can rewrite (using $p_1^1 = 1 - p_2^1, p_1^2 = 1 - p_2^2$) as

$$(+) \qquad w_2 - w_1 \geq t \cdot \frac{C_2 - C_1}{p_2^2 - p_2^1}.$$

The bonus contract $\tilde{w} = (\tilde{w}_1, \tilde{w}_2) = \left(0, t \cdot \frac{C_2 - C_1}{p_S^2 - p_S^1}\right)$ satisfies the IC condition (+) as an identity. We have $\overline{\tilde{w}}^2 = p_2^2 \tilde{w}_2$. If a different contract, say $w' = (w_1', w_2')$, also induces effort 2, then it must also satisfy (+). So we must have $w_1' \geq 0, w_2' \geq \tilde{w}_2$, and hence w' must cost at least as much as $\tilde{w}$, i.e., $\overline{w'}^2 = p_1^2 w_1' + p_2^2 w_2' \geq p_2^2 \tilde{w}_2$. It follows that $\tilde{w}$ optimally induces effort 2, as claimed.

[19] Sharecropping is a form of fixed-share contract, widely used in agrarian economies. Laffont and Matoussi (1995) investigate, theoretically and empirically, whether increasing the share in such contracts raises surplus. In Marschak and Wei (2019), we studys a similar problem for fixed-share contracts when there is no uncertainty but the set of efforts is allowed to be a continuum. Like a bonus contract, a fixed-share contract is linear in the revenues. The merits of linearity play a prominent role in the moral-hazard literature; see, e.g., Kim and Wang (1998), Bose *et al* (2011), Carroll (2015).

Given r, the Agent's net gain for an effort e is $r\overline{R}^e - tC_e$. Let $\hat{e}(r,t)$ denote the Agent's best response to r. It is the largest element of the set

$$\text{argmax}_{e\in\{1,\ldots,E\}}\left[r\overline{R}^e - tC_e\right].$$

The Principal keeps the fraction $1-r$ of expected revenue and the Agent receives the rest. If the Agent receives *all* of the expected revenue (i.e., $r=1$), then the effort he chooses maximizes welfare. Knowing the Agent's best response to every r, the Principal chooses the share $r^*(t)$, the smallest element of the set

$$\text{argmax}_{\ r\in[0,1]}\ [(1-r)\cdot\overline{R}^{\hat{e}(r,t)}].$$

We shall confine attention to triples $(\{C_e\}_{e=1,\ldots,E}\ ,\ \{R_s\}_{s=1,\ldots,S}\ ,\ \{p^e\}_{e=1,\ldots,E})$ for which

- $\hat{e}(r,t)$ exists for every pair (r,t) with $0<r\leq 1, t\geq 0$.
- The share $r^*(t)$ exists for every $t\geq 0$.

With an abuse of notation, we now let $\delta(t)$ denote the *Principal-favorite effort at t under fixed-share contracts*. We have $\delta(t)=\hat{e}(r^*(t),t)$.

We now have a counterpart of Proposition 1 for fixed-share contracts.

Proposition 2

The behavior of the Decentralization Penalty when fixed-share contracts are used is the same as the behavior (described in Proposition 1) when bonus contracts are used.

The proof of Proposition 2 exploits the same properties of γ and δ as does Proposition 1. We show that $\delta(\cdot)\leq\gamma(\cdot)$ (no squandering), that $\delta(\cdot)$ and $\gamma(\cdot)$ are weakly decreasing, and that $\delta(t)=\gamma(t)$ for sufficiently large and sufficiently small t. So the step functions $\gamma(\cdot),\delta(\cdot)$ have the same properties as in Lemmas 3 and 4, and so does the behavior of the Penalty.

6. Variations

The previous analysis concerned the welfare loss due to decentralization when contracts are restricted to certain types. In this section we show by an example that some of our results may no longer hold when contracts are unrestricted.[20] We also discuss an alternative criterion for judging the cost of decentralization, namely the loss in expected profit.

6.1 Unrestricted Contracts

If we do not restrict the Principal's contracts, the Penalty's behavior may be different. In particular, squandering may occur and there may be intervals where the Penalty continuously rises and other intervals where it continuously falls. That is illustrated in the following example. This example shows, moreover, that it may cost the Principal more to induce a lower effort.

Example 2

There are three efforts and three revenues ($E=S=3$). The revenue distributions for the three efforts are as follows:

[20] Recall from Section 4 that the restriction to bonus contracts is without loss if the IMCP condition holds. Thus the example we provide will have to violate that condition.

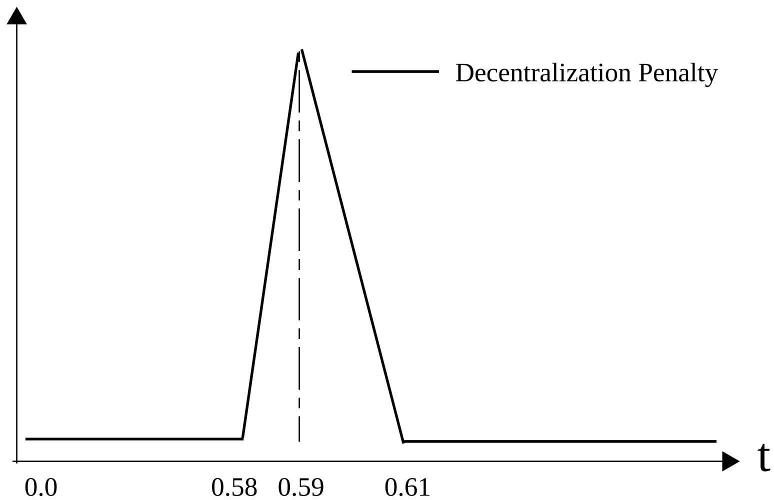

Figure 2 (Example 2)

	R_1	R_2	R_3
effort 1	$p_1^1 = \frac{5}{10}$	$p_2^1 = \frac{3}{10}$	$p_3^1 = \frac{2}{10}$
effort 2	$p_1^2 = \frac{4}{10}$	$p_2^2 = \frac{3}{10}$	$p_3^2 = \frac{3}{10}$
effort 3	$p_1^3 = \frac{1}{10}$	$p_2^3 = \frac{1}{10}$	$p_3^3 = \frac{8}{10}$

Let $(C_1, C_2, C_3) = (2, 4, 8)$ and let $(R_1, R_2, R_3) = (2, 3, 4)$.

Note that MLR is satisfied, but IMCP is violated.[21] The Penalty is graphed in Figure 2.[22]

[FIGURE 2 HERE]

We find that for $t < .58$ and $t \geq .61$ we have $\delta(t) = \gamma(t)$, so the Penalty is zero. For $.59 < t \leq .61$ we have $\delta(t) < \gamma(t)$. The Principal does not squander and the Penalty rises when t drops. But for $.58 < t \leq .59$ we have $\delta(t) > \gamma(t)$. The Principal squanders and the Penalty falls when t drops. A small variation of the example illustrates another consequence of violating IMCP: a higher effort may cost less to induce.[23]

6.2 An Alternative Criterion: Expected Profit

We now compare the Centralized and Decentralized modes in a new way. We judge them from the viewpoint of the firm's owner, who is concerned with expected profit, not welfare. The firm's expected revenue again depends on the effort the Agent chooses. In the Centralized mode, perfect monitoring enables the owner to enforce any effort she wants, and the Agent only needs to be compensated with the effort cost he incurs. Thus in the Centralized mode, the owner selects the effort $\gamma(t)$, a maximizer of expected surplus. In the Decentralized mode, there is no monitoring. The owner of the firm becomes a Principal, who generally has to pay the Agent more than the cost of his effort. The Principal will induce her favorite effort $\delta(t)$, as given in Definition 3.

While the decreasing step functions $\gamma(\cdot)$ and $\delta(\cdot)$ are the same as in our welfare analysis, the new Decentralization Penalty differs from the previous one. The expected-profit Penalty is

$$(4) \qquad \tilde{D}(t) \equiv \left[\overline{R}^{\gamma(t)} - tC_{\gamma(t)}\right] - \left[\overline{R}^{\delta(t)} - A_{\delta(t)}(t)\right] = \overline{R}^{\gamma(t)} - \overline{R}^{\delta(t)} + (A_{\delta(t)}(t) - tC_{\gamma(t)}),$$

where $A_e(t)$ again denotes the lowest cost of inducing effort e.

[21] We have $v_1 = C_1/p_3^1 = 10$, $v_2 = (C_2 - C_1)/(p_3^2 - p_3^1) = 20$, and $v_3 = (C_3 - C_2)/(p_3^3 - p_3^2) = 8$.

[22] As in Figure 1, the Penalty is calculated for all t in $\left\{\frac{1}{100}, \frac{2}{100}, \cdots, \frac{99}{100}, 1\right\}$. The graph is not drawn to scale, so that the narrow interval $[.58, .61]$ can be more easily visualized.

[23] If we let $(C_1, C_2, C_3) = (2, 4, 10)$, then we find that e_2 is optimally induced by the contract $(0, 20, 20)$ which costs 12, but e_3 is optimally induced by the contract $(0, 0, 13.33)$ which costs 10.67. In this variation IMCP is violated. We have $v_1 = C_1/p_3^1 = 2/(2/10) = 10$, $v_2 = (C_2 - C_1)/(p_3^2 - p_3^1) = 2/(1/10) = 20$, $v_3 = (C_3 - C_2)/(p_3^3 - p_3^2) = 6/(5/10) = 12$. Note that if there are only two efforts, then it can never happen that a higher effort costs less. It is easily shown that it costs the Principal tC_1 to (optimally) induce effort 1. If effort 2 (the only other effort) is optimally induced by a wage vector $\tilde{w}$, then $\tilde{w}$ satisfies the IR requirement $\overline{\tilde{w}}^2 \geq tC_2$, where $\overline{\tilde{w}}^2$ is the cost of the higher effort. Since $C_2 > C_1$, it costs the Principal more to induce the higher effort.

How does $\tilde{D}$ behave? Note first that from equations (2) and (4), we have

$$\tilde{D}(t) - D(t) = A_{\delta(t)}(t) - tC_{\delta(t)} \geq 0.$$

That is nonnegative because of the IR requirement. So the welfare Penalty is bounded from above by the expected-profit Penalty. This implies that whenever the Decentralization mode is preferred by the firm's owner, it is also preferred from the welfare point of view.

Nonetheless, the no-squandering condition (implied by the restriction to bonus contracts or fixed-share contracts) is not enough to generate definitive statementss about the shape of $\tilde{D}$.[24] So a sharp characterization of the expected-profit Penalty would require further assumptions. We leave this as an interesting direction for future research.

7. Economic Examples.

Can we find classic conditions on costs and revenues which imply that the Penalty behaves in a certain way? In our previous paper[25] — where there is no uncertainty and the set of efforts is allowed to be a continuum — results of that kind turned out to be scarce. But there was one such result. To obtain it, we assumed that the effort set is an interval and the firm is a price-taker: each effort is a product quantity and the product sells at a price which the firm takes as given. So marginal revenue is flat. We also assumed that marginal cost increases linearly. We let the Principal use a fixed-share contract, where the share is the Principal's favorite. We then found that at every t the Penalty is decreasing in t: a technical improvement (a drop in t) smoothly raises the Penalty. Is there a counterpart of that result in the present paper's model?

In the present paper's model there are S revenues and E efforts. The "flat marginal revenue" condition is:

$$\text{there exists } H \text{ such that } R_s - R_{s-1} = H \text{ for all } s \text{ in } \{2, \ldots, S\}.$$

The "increasing marginal cost" condition is:

$$C_2 - C_1 < C_3 - C_2 < \cdots < C_{E-1} - C_{E-2} < C_E - C_{E-1}.$$

If bonus or fixed share contracts are used, then *whether or not* marginal revenue is flat and marginal cost is increasing, we have the Penalty behavior of Proposition 1. So if there is an interval where the Penalty smoothly changes when t drops, then it must be an interval where the Penalty increases when t drops. That is also true in the price-taking-firm result of our previous paper.

Our moral-hazard model, however, has a weakness: the probabilities are not explained. Can we obtain results about the Penalty if costs and revenues obey classic conditions and the probabilities are *endogenously* determined from the nature of the firm's effort-choosing task? Do those probabilities obey the MLR condition?

Consider the following example.

[24] Indeed, the behavior of $\tilde{D}$ depends on the sign of $A_{\delta(t)}(t) - tC_{\gamma(t)}$. Even if $\delta(t) \leq \gamma(t)$ for all t (so that $tC_{\gamma(t)} \geq tC_{\delta(t)}$), the comparison between $A_{\delta(t)}(t)$ and $tC_{\gamma(t)}$ is still ambiguous because IR implies that $A_{\delta(t)}(t) \geq tC_{\delta(t)}$.

[25] Marschak and Wei (2019).

The firm has E potential buyers of its product. A true buyer buys one unit, but only after personal contact with the firm. The true buyer pays the firm a price that equals one plus the cost of producing one unit. So the revenue collected by the firm equals the number of contacted buyers who turn out to be true buyers. Advertising effort, chosen by an Agent, increases the number of potential buyers who would turn out to be true buyers if they were contacted. The Agent's effort is induced by a Principal, who (as before) induces her favorite effort

The possible advertising efforts are $1, \ldots, e, \ldots E$. If the effort e is spent on advertising, then e of the E potential buyers become true buyers when contacted. Effort e costs tC_e, where C_e is strictly increasing in e. $S \leq E$ potential buyers are randomly chosen and contacted. If s of the S contacted buyers turn out to be true buyers then the firm collects the revenue R_s. So the possible revenues are $\{R_1, \ldots, R_s, \ldots, R_S\} = \{1, 2, \ldots, s, \ldots, S\}$. (Thus the "flat marginal revenue" condition is satisfied). The S randomly chosen potential buyers are a sample without replacement. For a given effort e, the probability that the sample contains s true buyers (so revenue is R_s) is given by the hypergeometric distribution. So

$$p_s^e = \frac{\binom{e}{s} \cdot \binom{E-e}{S-s}}{\binom{E}{S}},$$

and, in particular,

$$p_S^e = \frac{\binom{e}{S}}{\binom{E}{S}}.$$

The hypergeometric distribution has the MLR property.[26] Hence so does our family of probabilities $\{p_s^e\}_{e=1,\ldots,E;s=1,\ldots,S}$. If we now *require* the firm to use bonus (or fixed-share) contracts, then we get the Proposition-1 behavior of the Penalty.[27]

Note that a number of other distributions have the MLR property. Some of them may again be implied by the firm's effort-choosing task. Whether endogenous probabilities and classic economic conditions restrict the behavior of the Penalty is a question that merits further attention.[28]

8. Concluding Remarks

[26]That means that the fraction

$$\frac{\text{probability that the sample has } i \text{ true buyers when effort is } e}{\text{probability that the sample has } j \text{ true buyers when effort is } e}$$

is increasing in e if $i > j$. For a proof that the hypergeometric distribution has the MLR property see Lehman (1986), p.80.

[27]If the IMCP condition were satisfied, then the firm loses nothing if it uses bonus contracts, and we get the Penalty behavior of Proposition 1 without requiring the firm to use bonus contracts. Unfortunately, however, we can easily find $C_1, \ldots, C_E$ so that the IMCP condition is violated even though marginal cost is increasing and marginal revenue is flat. Let $E = 20, S = 10$. Then we find — rounding to five decimal places — that $p_S^1 = .00036, p_S^2 = .00542$. Now let $C_e = e^2$ for all e in $\{1, \ldots, E\}$, so the increasing-marginal-cost condition is satisfied. Thus $C_1 = 1, C_2 = 4$. IMCP requires that $v_1 \leq v_2$, where $v_1 = C_1/p_S^1, v_2 = (C_2 - C_1)/(p_S^2 - p_S^1)$. That is equivalent to $C_1 p_S^2 \leq C_2 p_S^1$ or $p_S^2 \leq 4p_S^1$. That is violated, since $.00542 > (4) \cdot (.00036) = .00144$.

[28]Economic examples where IMCP holds endogenously would be of particular interest. Note that if marginal cost is flat or rising, then we have IMCP if $p_S^e - p_S^{e-1}$ is decreasing in e.

The dramatic and rapid advances that we observe in production technology and in monitoring technology strongly motivate a better understanding of the effect of those improvements on a firm designer's choice between the Decentralized and Centralized modes. But it is difficult to formulate the question with sufficient precision so that answers can be found. Clearly we have to start with highly simplified models. It is natural to try the standard moral-hazard Principal/Agent framework — with a finite collection of production efforts, and uncertainty about the revenue they earn — as our model of the Decentralized mode.

We have obtained sharp predictions about the behavior of the Decentralization Penalty in response to improvements in production technology. We found that when the Principal is restricted to use either bonus contracts or fixed-share contracts, the Decentralization Penalty will oscillate in a continuous-rise-sudden-change cycle until production technology becomes sufficiently advanced. If we require expected revenues and costs to obey the Quasi-concavity condition, then the sudden changes are sudden drops. There are no intervals where technology improvement smoothly lowers the Penalty. These results use the fact that under both contract types the Principal never squanders. While an improvement in monitoring technology always strengthens the case for the Centralized mode, advances in production technology may do the opposite.

There are many ways to vary and extend our model. Here are some of them.

- Find other contract types where the Principal never squanders and study the Penalty for each type.
- Study the Penalty when the collection of possible efforts is a continuum. That may be less realistic, but it permits calculus-based techniques to be used.[29]
- Make the Agent risk-averse. In the simplest model, the Agent chooses the effort x_e if $\overline{w^*} \geq tC_e$ and $\overline{w^*} - tC_e \geq \overline{w^*} - tC_f$ for all f, where $w^* = (u(w_1), \ldots, u(w_S))$ and u is a strictly concave function. There may be functions u which allow us to study the behavior of the Penalty.
- Let t be a random variable whose average drops when technology improves. Let t be observed by only one of the two parties.
- Let there be several Agents. In the easiest case there are two Agents and the parameter t is known to all three parties. Realized revenue depends on the two Agents' efforts, and the probability of a given revenue for a given effort pair is common knowledge. The Principal chooses two shares whose sum must lie between zero and one. Then in the Decentralized mode we have a three-player game for every t. Each Agent chooses an effort and the Principal chooses the two shares. Suppose that for every t the game has a unique pure-strategy equilibrium. To compute the Decentralization Penalty we first find the effort pair that maximizes surplus. We compare that surplus with surplus at the equilibrium. When t drops, does the Penalty rise or fall?

[29] A continuum model is studied in Nasri, Bastin, and Marcotte (2015). There the set of possible effort levels is an interval, but the set of possible revenues remains finite. The Principal's optimally-induce-x_e problem becomes a "semi-infinite" linear programming problem. (For a survey of semi-infinite linear programming, see Goberna, M. and López, 2002.)

These variations, and others, deserve attention.

APPENDIX

Proof of Lemma 1

It suffices to show that an effort which maximizes the Principal's net gain at t^* cannot be less then an effort which maximizes it at $t^{**} < t^*$. That is implied by the following stronger statement, which we now prove:

(A1) If $e > f$ and $\overline{R}^e - A_e(t^*) \geq \overline{R}^f - A_f(t^*)$, then for $t^{**} < t^*$ we have $\overline{R}^e - A_e(t^{**}) \geq \overline{R}^f - A_f(t^{**})$.

At every t, the Principal's (primal) optimally-induce-e problem is the following.

Find nonnegative wages $w_1, \ldots, w_S$ which minimize $p_1^e w_1 + \cdots + p_S^e w_S$ subject to:

$$\begin{array}{lccccl}
p_1^e w_1 + p_2^e w_2 & + \cdots + & p_S^e w_S & \geq & tC_e & \boxed{\text{IR}} \\
(p_1^e - p_1^1) \cdot w_1 + (p_2^e - p_2^1) \cdot w_2 & + \cdots + & (p_S^e - p_S^1) \cdot w_S & \geq & t \cdot (C_e - C_1) & \boxed{e \text{ is not worse than effort } 1} \\
(p_1^e - p_1^2) \cdot w_1 + (p_2^e - p_2^2) \cdot w_2 & + \cdots + & (p_S^e - p_S^2) \cdot w_S & \geq & t \cdot (C_e - C_2) & \boxed{e \text{ is not worse than effort } 2} \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
(p_1^e - p_1^E) \cdot w_1 + (p_2^e - p_2^E) \cdot w_2 & + \cdots + & (p_S^e - p_S^E) \cdot w_S & \geq & t \cdot (C_e - C_E) & \boxed{e \text{ is not worse than effort } E}
\end{array}$$

Recall that we confine attention to situations where the primal problem has a solution at every t. We now state the dual of the preceding problem. We let h denote the dual variable ("shadow price") associated with the IR constraint and we let $y_1, y_2, \ldots, y_E$ denote the dual variables associated with the E IC constraints. Then the dual becomes the following.

Find nonnegative "shadow prices" $h, y_1, \ldots, y_E$ which maximize

$$t \cdot [hC_e + y_1 \cdot (C_e - C_1) + y_2 \cdot (C_e - C_2) + \cdots + y_E \cdot (C_e - C_E)]$$

subject to:

$$\begin{array}{ccccccccccc}
h & + & y_1 \cdot (p_1^e - p_1^1) & + & y_2 \cdot (p_1^e - p_1^2) & + & \cdots & + & y_E \cdot (p_1^e - p_1^E) & \leq & p_1^e \\
h & + & y_1 \cdot (p_2^e - p_2^1) & + & y_2 \cdot (p_2^e - p_2^2) & + & \cdots & + & y_E \cdot (p_2^e - p^E s_2) & \leq & p_2^e \\
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
h & + & y_1 \cdot (p_S^e - p_S^1) & + & y_2 \cdot (p_S^e - p_S^2) & + & \cdots & + & y_E \cdot (p_S^e - p_S^E) & \leq & p_S^e
\end{array}$$

Strong duality tells us: (1) since the primal has a solution, say $w = (w_1, \ldots, w_S)$, the dual also has a solution, say $(h, y_1, \ldots, y_E)$; (2) the value of the minimand in a solution to the primal equals the value of the maximand in a solution of the dual. That means that inducing e costs the Principal

$$A_e(t) = \overline{w}^e = tJ_e,$$

where

$$J_e = hC_e + y_1 \cdot (C_e - C_1) + y_2 \cdot (C_e - C_2) + \cdots + y_E \cdot (C_e - C_E).$$

We have $J_e > 0$, since (by the IR requirement) $A_e(t) \geq tC_e > 0$. Consider $f > e$ and the optimally-induce-f problem. (By assumption, that problem has a solution at every t). The Principal (weakly) prefers f to e at t if

$$(A2) \qquad \overline{R}^f - \overline{R}^e \geq t \cdot (J_f - J_e)$$

(since $A_e(t) = tJ_e, A_f(t) = tJ_f$). Suppose $J_f \geq J_e$ and apply (A2) to the case $t = t^*$ and the case $t = t^{**}$. We see that if the Principal (weakly) prefers f to e at $t = t^*$, then she continues to do so at $t = t^{**} < t^*$. Suppose, on the other hand, that $J_f < J_e$ and apply (A2) again. The left side is positive; that follows from the MLR assumption. The right side is negative. So (A2) holds for $t = t^*$ and for $t = t^{**} < t^*$.

Thus the Principal cannot switch from weakly preferring f at $t = t^*$ to strongly preferring e at $t = t^{**} < t^*$. That establishes (A1), which implies Proposition A. □

Proof of Lemma 3

The proof is analogous to the proof of Lemma 1. It suffices to show that an effort which maximizes the Principal's net gain at t^*, when she uses bonus contracts, cannot be less then an effort which maximizes it at $t^{**} < t^*$. That is implied by the following stronger statement, which we now prove:

$(A3)$ If $e > f$ and $\overline{R}^e - A^b_e(t^*) \geq \overline{R}^f - A^b_f(t^*)$, then for $t^{**} < t^*$ we have $\overline{R}^e - A^b_e(t^{**}) \geq \overline{R}^f - A^b_f(t^{**})$.

The Principal's (primal) optimally-bonus-induce-e problem is:

Find a nonnegative z which minimizes $p^e_S z$ subject to:

$$\begin{array}{rcl} z & \geq & tC_e \\ z \cdot (p^e_S - p^1_S) & \geq & t \cdot (C_e - C_1) \\ z \cdot (p^e_S - p^2_S) & \geq & t \cdot (C_e - C_2) \\ \vdots & & \vdots \\ z \cdot (p^e_S - p^E_S) & \geq & t \cdot (C_e - C_E) \end{array}$$

The dual of this minimization problem is:

Find nonnegative "shadow prices" $h, y_1, \ldots, y_E$ which maximize

$$t \cdot [hC_e + y_1 \cdot (C_e - C_1) + y_2 \cdot (C_e - C_2) + \cdots + y_E \cdot (C_e - C_m)]$$

subject to:

$$hC_e + y_1 \cdot (p^e_S - p^1_S) + y_2 \cdot (p^e_S - p^2_S) + \cdots + y_E \cdot (p^e_S - p^E_S).$$

We confine attention to situations where the primal problem has a solution, say $z^e(t)$, which costs the Principal $A^b_e(t) = p^e_S z^e(t)$. Since $C_e > 0$ and $z_e(t)$ satisfies the IR condition $z_e(t) \geq tC_e$, where $C_e > 0$, we have $z_e(t) > 0$ and $A^b_e(t) > 0$. By strong duality: (1) the dual also has a solution, say $(h, y_1, \ldots, y_E)$, and (2) the value of the minimand in the solution to the primal equals the value of the maximand in the solution to the dual. That means

$$A^b_e(t) = t\tilde{J}_e, \text{ where } \tilde{J}_e = hC_e + y_1 \cdot (p^e_S - p^1_S) + \cdots + y_E \cdot (p^e_S - p^E_S) \text{ and } \tilde{J}_e > 0.$$

That holds as well when we replace "e" with "f". We now verify (A3) by repeating the argument used to verify (A1) in the proof of Proposition A, with $\tilde{J}_e, \tilde{J}_f$ replacing J_e, J_f. That completes the proof. □

Proof of Lemma 4

We first show, for a fixed t, that if $f > e$, then the Agent (weakly) prefers f to e: the Agent's net gain when f is optimally bonus-induced is at least as large as the net gain when e is optimally bonus-induced, i.e.

$$\text{(A4)} \qquad [p^f_S \cdot z^f(t) - tC_f] - [p^e_S \cdot z^e(t) - tC_e] \geq 0 \text{ if } f > e.$$

That is the case since $z^f(t)$ satisfies the "f-is-better-than-e" requirement

$$p^f_S z^f(t) - tC_f \geq p^e_S z^f(t) - tC_e.$$

Using the fact that $z^f(t) > z^e(t)$ whenever $f > e$ (established in Section 4.1), we obtain

$$\text{(A5)} \qquad p^e_S z^f(t) - tC_e \geq p^e_S z^e(t) - tC_f \text{ whenever } f > e.$$

Next we show that the Principal strictly prefers optimally bonus-inducing the largest welfare-maximizing effort to optimally bonus-inducing any effort which is larger than that.[30] That implies no squandering: the Principal's favorite effort $\delta(t)$ cannot exceed $\gamma(t)$. (If it did, there would be at least one effort (namely $\gamma(t)$) which the Principal prefers to $\delta(t)$). We claim that

$$\text{(A6)} \qquad [\overline{R}^{\gamma(t)} - p^{\gamma(t)}_S \cdot z^{\gamma(t)}] - [\overline{R}^d - p^d_S \cdot z^d(t)] > 0 \text{ for all } d > \gamma(t).$$

To show this we note that

$$\begin{aligned} [\overline{R}^{\gamma(t)} - p^{\gamma(t)}_S \cdot z^{\gamma(t)}] - [\overline{R}^d - p^d_S \cdot z^d(t)] &= \underbrace{[\overline{R}^{\gamma(t)} - tC_{\gamma(t)}] - [\overline{R}^d - tC_d]}_{V} \\ &+ \underbrace{[p^d_S \cdot z^d(t) - tC_d]}_{W} - \underbrace{[p^{\gamma(t)}_S \cdot z^{\gamma(t)} - tC_{\gamma(t)}]}_{Y} \end{aligned}$$

[30]The argument is an adaptation of Step 1 in the proof of Theorem 1 in Balmaceda *et al* (2016).

Since $\gamma(t)$ is the largest welfare-maximizing effort, we have $V > 0$. Applying (A5), we have $W \geq Y$. So (A6) is established. That completes the proof. $\square$

Proof of Proposition 1

Recall that

$$D(t) = [\overline{R}^{\gamma(t)} - \overline{R}^{\delta(t)}] + t \cdot \left(C_{\delta(t)} - C_{\gamma(t)}\right).$$

First we show that there exists $\bar{T} > 0$ such that $\gamma(t) = \delta(t) = 1$ for all $t > \bar{T}$. By MLR, we know that $\overline{R}^e - \overline{R}^1 > 0$ for all $e > 1$. Therefore, there exists $\bar{T}_1 > 0$ such that

$$\overline{R}^e - \overline{R}^1 < t(C_e - C_1), \forall e > 1, t > \bar{T}_1.$$

So $\gamma(t) = 1$ for all $t > \bar{T}_1$. We now turn to the Principal-favorite effort $\delta(t)$. We saw in the proof of Lemma 3 that $A^b_e(t) = \tilde{J}_e t$ for some $\tilde{J}_e \geq C_e$. Moreover, it is easy to verify that effort 1 can be optimally bonus-induced by the wage vector $(0, ..., 0, tC_1/p^1_S)$, and so we have $\tilde{J}_1 = C_1 < C_e \leq \tilde{J}_e$, for all $e > 1$. Therefore there exists $\bar{T}_2 > 0$ such that

$$\overline{R}^e - \overline{R}^1 < t(\tilde{J}_e - \tilde{J}_1), \forall e > 1, t > \bar{T}_2.$$

So $\delta(t) = 1$ for all $t > \bar{T}_2$. Letting $\bar{T} = \max\{\bar{T}_1, \bar{T}_2\}$, we conclude that $\gamma(t) = \delta(t) = 1$ for all $t > \bar{T}$.

Next we show that there exists $\bar{t} > 0$ such that $\gamma(t) = \delta(t) = E$ for all $t \in (0, \bar{t})$. By MLR, we know that $\overline{R}^E - \overline{R}^e > 0$ for all $e < E$. Therefore there exists $\bar{t}_1$ such that

$$\overline{R}^E - \overline{R}^e > t(C_E - C_e), \forall e < E, t < \bar{t}_1.$$

So $\gamma(t) = E$ for all $t \in (0, \bar{t}_1)$. Moreover, there exists $\bar{t}_2 > 0$ such that

$$\overline{R}^E - \overline{R}^e > t(\tilde{J}_E - \tilde{J}_e), \forall e < E, t < \bar{t}_2.$$

So $\delta(t) = E$ for all $t \in (0, \bar{t}_2)$. Letting $\bar{t} = \min\{\bar{t}_1, \bar{t}_2\}$, we conclude that $\gamma(t) = \delta(t) = E$ for all $t \in (0, \bar{t})$.

We cannot have $\bar{T} < \bar{t}$.[31] If $\bar{T} = \bar{t}$, then $D(t) = 0$ at all $t > 0$. So we now turn to the case where there are intermediate values of t, i.e., $\bar{T} > \bar{t}$.

Recall that any interval on which the weakly decreasing and left-continuous step functions $\gamma(\cdot)$ and $\delta(\cdot)$ are decreasing and are not equal has a subinterval where each is constant. So we can partition the interval $(\bar{t}, \bar{T})$ into non-overlapping left-open subintervals, defined by numbers $t_1, ..., t_Q$, where $\bar{t} = t_1 < t_2 < ... < t_Q = \bar{T}$. The subintervals are

$$(t_1, t_2], (t_2, t_3], \ldots, (t_{q-1}, t_q], \ldots, (t_{Q-1}, t_Q].$$

Since, by Lemma 4, we have $\gamma(\cdot) \geq \delta(\cdot)$, we can choose the subintervals so that each has the following property:

[31] We have (1) $\gamma(t) = \delta(t) = 1$ for all $t > \bar{T}$; and (2) $\gamma(t) = \delta(t) = E$ for all $t \in (0, \bar{t})$. If we had $\bar{T} < \bar{t}$, then there would exist $\hat{t} \in (\bar{T}, \bar{t})$ such that we have both $\gamma(\hat{t}) = \delta(\hat{t}) = 1$ and $\gamma(\hat{t}) = \delta(\hat{t}) = E$, which is a contradiction since $E > 1$.

Either $\gamma(\cdot) = \delta(\cdot)$ and hence $D(\cdot) = 0$ throughout the subinterval; or, for some γ', δ' with $\gamma' > \delta'$, we have $\gamma(\cdot) = \gamma', \delta(\cdot) = \delta'$ throughout the subinterval and hence

$$D(t) = \overline{R}^{\gamma'} - \overline{R}^{\delta'} + t \cdot (C_{\gamma'} - C_{\delta'}) > 0$$

throughout the subinterval.

Next, it will be convenient to have a new symbol for surplus at effort e for a given t. Define

$$\Omega(e, t) := \overline{R}^{e} - tC_e.$$

Recall that $\gamma(t)$ denotes the smallest maximizer of $\Omega(e, t)$. If quasi-concavity is satisfied then

(A7) $\quad e' < e'' < \gamma(t)$ implies $\Omega(\gamma(t), t) - \Omega(e', t) < \Omega(\gamma(t), t) - \Omega(e'', t).$

Informally: welfare drops whenever effort moves further below the lowest welfare-maximizing effort. Note that

$$D(t) = \Omega(\gamma(t), t) - \Omega(\delta(t), t).$$

Consider the interval $(t_q, t_{q+1}]$, where $1 \leq q \leq Q - 1$. We shall now see that quasi-concavity rules out a downward "jump" in $D(\cdot)$ just after t_{q+1}. Formally, it is *NOT* the case that

(A8) $\quad D(t_{q+1}) > 0$ and for all sufficiently small positive ϵ we have $D(t_{q+1}) > D(t_{q+1} + \epsilon)$.

The first step is to note, using the maximum theorem, that since, for a fixed t, $\gamma(t)$ maximizes $\Omega(e, t)$, the function $\Omega(\gamma(t), t)$ is continuous in t. So if we indeed have the discontinuity at t_{q+1} that is described in (A8), then the left-continuous function $\Omega(\delta(t), t)$ must exhibit a discontinuity at t_{q+1}. By Lemma 1, $\delta(\cdot)$ is (weakly) decreasing in t. By Lemma 4, $\delta(\cdot) \leq \gamma(\cdot)$. So quasi-concavity implies that an increase in t moves welfare further below its maximum. Hence the only possible discontinuity in the function $\Omega(\delta(\cdot), \cdot)$ at t_{q+1} is a "downward" jump just after t_{q+1}.

Recall that $\delta(\cdot)$ is weakly decreasing in t. We have such a downward jump if there exists $L > 0$ such that

(A9) $\quad \Omega(\delta(t_{q+1}), t_{q+1}) - \Omega(\delta(t_{q+1} + \epsilon), t_{q+1} + \epsilon) \geq L$ for all sufficiently small positive ϵ.

Now note that (A8) can be rewritten:

for all sufficiently small positive ϵ we have

$$\Omega(\gamma(t_{q+1}), t_{q+1}) - \Omega(\gamma(t_{q+1} + \epsilon), t_{q+1} + \epsilon) > \Omega(\delta(t_{q+1}), t_{q+1}) - \Omega(\delta(t_{q+1} + \epsilon), t_{q+1} + \epsilon).$$

Since $\Omega(\gamma(\cdot), \cdot)$ is continuous we can make the term on the left of that inequality as small as we wish by choosing a sufficiently small positive ϵ. But, by (A9), the term on the right is at least $L > 0$ whatever ϵ may be. So we have a contradiction to (A8). The downward jump just after t_{q+1}, described in (A8), cannot occur.

There may, however be an upward jump at t_{q+1}. That happens if

$$\text{for all sufficiently small positive } \epsilon \text{ we have } D(t_{q+1}) < D(t_{q+1} + \epsilon).$$

Now let G_{q+1} denote $\inf\{D(t) \ : \ t \in (t_q, t_{q+1}]\}$ and let u_q denote $D(t_{q+1})$. If there is an upward jump just after t_{q+1}, then we have $u_q < G_{q+1}$. If $D = 0$ throughout the interval $(t_q, t_{q+1}]$ then $u_q = G_{q+1}$.

So we have exhibited the pairs (G_q, u_q) described in the Proposition. That concludes the proof. $\square$

We noted in the text that if we drop the Quasi-concavity assumption (but retain MLR) then we cannot exclude the possibility that when technology improves (t drops) there is an upward jump at some t. To be specific, it is no longer true that the only possible discontinuity in the function $\Omega(\delta(\cdot), \cdot)$ (when we move from left to right) is a "downward" jump. That means that (A8) can now be satisfied.

Proof of Proposition 2

We shall establish fixed-share analogs of Lemmas 3 and 4. We then apply them to prove the proposition.

Proving that t^*, t^{**} implies $\delta(t^{**}) > \delta(t^*)$.

We use a standard proposition from monotone comparative statics. (See, for example, Sundaram (1996)).

Consider sets $U \in \mathbb{R}, V \in \mathbb{R}$ and a function $h : U \times V \rightarrow \mathbb{R}$. The two arguments of h are denoted u, v. The function h displays *strictly increasing differences in the variables* u, v if

$$h(u_H, v_H) - h(u_L, v_H) > h(u_H, v_L) - h(u_L, v_L)$$

whenever $u_H, u_L \in U$, $v_H, v_L \in V$, $u_H > u_L$, and $v_H > v_L$. We use the following proposition:

(#) Suppose that for every $v \in V$, the problem

$$\text{maximize } h(u, v) \text{ subject to } u \in U$$

has at least one solution. Suppose also that h satisfies strictly increasing differences in u, v. Consider $v_H, v_L \in V$ with $v_H > v_L$. Let u_H be a maximizer of $h(u, v_H)$ on U and let u_L be a maximizer of $h(u, v_L)$ on U. Then $u_H \geq u_L$.

Note the following:

(α) If h takes the form $h(u, v) = f(u, v) + g(u)$, then h displays strictly increasing differences in u, v if and only if f displays strictly increasing differences in u, v.

(β) If h takes the form $h(u, v) = f(u) \cdot g(v)$ and f and g are strictly increasing, then h displays strictly increasing differences in u, v.

It suffices to consider any effort that maximizes the Principal's net gain at $t = t_H$ and any effort that maximizes the Principal's net gain at $t = t_L < t_H$, and to show that the first of these efforts cannot be less than the second. So it suffices to prove

$$(A10) \qquad \hat{e}(r^*(t_L), t_L) \geq \hat{e}(r^*(t_H), t_H) \text{ whenever } 0 < t_L < t_H.$$

We note first that the Agent's chosen effort $\hat{e}(r,t)$ depends only on the ratio $\frac{r}{t}$, which we shall call ρ. The set of possible values of ρ is $\left(0, \frac{1}{t}\right]$. The Agent's effort is a value of e which maximizes $r\overline{R}^e - tC_e = t \cdot (\rho\overline{R}^e - C_e)$ and is therefor a maximizer of $\rho\overline{R}^e - C_e$. We shall use a new symbol, namely $\phi(\rho)$ to denote the Agent's chosen effort when the ratio is ρ. So $\phi(\rho) = \hat{e}(r,t)$. In view of (α),(β), and the fact that $\overline{R}^e$ is nondecreasing in e, the function $\rho\overline{R}^e - C_e$ displays strictly increasing differences with respect to ρ, e. Hence (applying $(\#)$) the maximizer $\phi(\rho)$ is nondecreasing in ρ, so we have

$$(A11) \qquad \phi(\rho_H) \geq \phi(\rho_L) \text{ whenever } 0 < \rho_L < \rho_H.$$

We can now reinterpret the Principal as the chooser of a ratio. For a given t, he chooses the ratio $\rho^*(t) = \frac{r^*(t)}{t}$, where

$$\rho^*(t) = \min\{\text{argmax}_{\rho \in (0,1/t)}\ M(\rho, -t)\ \},$$

and

$$M(\rho, -t) = (1 - t\rho) \cdot \overline{R}^{\phi(\rho)} = \overline{R}^{\phi(\rho)} - t\rho\overline{R}^{\phi(\rho)}.$$

By (α), the function M has strictly increasing differences in $\rho, -t$ if the function $-t\rho\overline{R}^{\phi(\rho)}$ has strictly increasing differences in $\rho, -t$. Examining $[-t] \cdot \left[\rho \cdot \overline{R}^{\phi(\rho)}\right]$, we see that the first expression in square brackets is strictly increasing in $-t$. The second expression is strictly increasing in ρ, since $\overline{R}^e$ is nondecreasing in e and, by (A7), ϕ is nondecreasing. So we can apply (β). The function $-t\rho\overline{R}^{\phi(\rho)}$ indeed has strictly increasing differences in $\rho, -t$, and so, since $\rho^*(t)$ is a maximizer of $M(\rho, -t)$,

$$(A12) \qquad \frac{r^*(t_L)}{t_L} = \rho^*(t_L) \geq \rho^*(t_H) = \frac{r^*(t_H)}{t_H} \text{ whenever } 0 < t_L < t_H.$$

But $\phi\left(\frac{r^*(t)}{t}\right) = \hat{e}(r^*(t), t)$. That, together with (A11) and (A12), establishes (A10).

Proving that $\delta(\cdot) \leq \gamma(\cdot)$ (no squandering)

In view of (α), (β), the function $r \cdot \overline{R}^e - tC_e$, where t is fixed, displays strictly increasing differences in r and e, since $\overline{R}^e$ is nondecreasing in e. Since, for fixed t, the effort $\hat{e}(r,t)$ maximizes $r \cdot \overline{R}^e - tC_e$, we obtain

$$\hat{e}(r_L, t) \leq \hat{e}(r_H, t) \text{ whenever } 0 < r_L < r_H \leq 1.$$

In particular (since $r^*(t) \leq 1$),

$$\hat{e}(r^*(t), t) \leq \hat{e}(1, t) \text{ for all } t > 0.$$

That implies that $\delta(t) \leq \gamma(t)$. (The Principal never squanders).

Proving that the conditions which implied Proposition 1 hold again

It is straightforward to show that for sufficiently small t and any $r \in (0,1]$ we have $\hat{e}(r,t) = Es$ and for sufficiently large t and any $r \in (0,1]$ we have $\hat{e}(r,t) = 1$. That implies that we have the same statement as we had in the proof of Proposition 1: the Penalty is zero for sufficiently small and sufficiently large t.

We have shown that the step functions $\gamma(\cdot), \delta(\cdot)$ (specifying the largest welfare-maximizing effort, and the Principal's favorite effort for fixed-share contracts) have the same properties as in Proposition 1. So for fixed-share contracts the Penalty indeed behaves as in Proposition 1.

That completes the proof. □

REFERENCES

Acemoglu, D., Wolitzky, A., (2011). "The economics of labor coercion". *Econometrica*, **79**, 555-600.

Balbaieff M., M. Feldman, N. Nisan (2009). "Free riding and free labor in combinatorial agency". In Mavronicolas and Papadopoulou (eds.) SDAGT 2009. LNCS, vol. 5814, 109-121. Springer.

Ball, K. (2010). "Workplace surveillance: an overview", *Labor History*, **51**, 897-106.

Balmaceda, F., S. Balseiro, J. Correa, N. Stier-Moses, (2016). "Bounds on the welfare loss from moral hazard with limited liability". *Games and Economic Behavior*, **95**, 137-155.

Fong, Y. and Li, J (2016). "Information revelation in relational contracts", *Rreview of Economic Studies*, **83**, 1-27.

Goberna, M. and López, M, (2002). "Linear semi-infinite programming theory: an updated survey",*European Journal of Operations Research*, **143**, 390-405.

Grossman, S. and Hart, O. (1983). "An analysis of the principal-agent problem". *Econometrica*, **51**, 7-45.

Herweg F., Müller, D., Weinschenk, P. (2010). "Binary payment schemes: moral hazard and loss aversion". *American Economic Review*, **100**, 2451-2477.

Holmstrom, B. (1979). "Moral hazard and observability", *The Bell Journal of Economics*, **10**, 74-91.

Holmstrom, B. (1982). "Moral hazard in teams", *The Bell Journal of Economics*, **13**, 324-340.

Lehman, E. (1986). *Testing Statistical Hypotheses, Second Edition*, John Wiley and Sons.

Marschak, T. and Wei, D. (2019). "Technical change and the decentralization penalty", in W. Trockel, ed., *Social design: Essays in Memory of Leonid Hurwicz*, Springer.

Mateescu, A., Nguyen, A. (2019). “Workplace monitoring and surveillance”, *Data and Society*, February.

Moulin, H. (2006). “The price of anarchy of serial, average, and incremental cost sharing”. *Economic Theory*, **36**, 379-405.

Nasri,M., F. Bastin, P. Marcotte (2015). “Quantifying the social welfare loss in moral hazard models”. *European Journal of Operations Research*, **245**, 226-235.

Nissan,N., T. Roughgarden, E. Tardos, V. Vazirani (2007). *Algorithmic Game Theory*, Cambridge University Press.

Ross, S. (1973). “The economic theory of agency: the principal’s problem”. *American Economic Review*, **63**.

Roughgarden, T. (2005). *Selfish Routing and the Price of Anarchy*, MIT Press.

Sundaram, R.K. (1996). *A First Course in Optimization Theory*, Cambridge University Press.

Schmitz, P. (2005) “Workplace surveillance, privacy protection, and efficiency wages”. *Labour Economics*, **12**, 727-738.

Tippet, E., Alexander, C., Eigen, Z.(2018). “When timekeeper software undermines compliance”, *Yale Journal of Law and Technology*, **19**, Issue 1.

Vagle, S. (2020). “Cybersecurity and moral hazard”, *Stanford Technology Law Review*, **23**, 102-113.

Zuboff, S. (2019) *The Age of Surveillance Capitalism*, Public Affairs.

Acknowlegement:

We are grateful to Ronnie Miller, who provided excellent research assistance. That included building the codes which discovered Example 1 and Example 2.