Analysis of Hidden Gaming in a Three-Level Hierarchy

JEAN-JACQUES LAFFONT
GREMAQ, Toulouse

1. INTRODUCTION

The principal agent model is a convenient tool for analyzing the internal structure of organizations. The design of optimal contracts by a principal to mitigate the moral hazard or adverse selection problems embedded in his relationship with one agent (or several agents of the same type) is now well-understood. However, this work is restricted to two-level hierarchies.

The economic analysis of more complex organizations is still very limited. There exists an old literature studying the optimal design of communication channels, leaving aside incentive questions. Recent work includes Crémer (1979), Geanakoplos and Milgrom on hierarchical structures, Sah and Stiglitz, Green and Laffont (1986), and Marschak and Reichelstein on the comparison of hierarchical structures and more general networks. Multilayers of principal agent relationships with pure moral hazard have been considered (Williamson, Mirrlees, Calvo and Wellisz). In this work, as in Stiglitz, which is concerned with adverse selection, coalitional behavior was not considered...

I am grateful to the Department of Economics of the Australian National University for its hospitality. A first draft of this paper was written while preparing an invited lecture at the Bicentennial Australian Economics Congress in Canberra. I benefited from the reactions of Patrick Bolton, Jean Tirole, Patrick Rey, and Robert Wilson to an earlier draft, and from the comments of the referees of this journal.

1. See the synthetic treatments by Hart and Holmstrom for moral hazard and Guensterie and Laffont for adverse selection.
and supervisors, when introduced, were assumed to be benevolent or reliant on hard information that could not be manipulated.

The analysis of soft (or manipulable) information within organizations has been considered by Antle and more broadly by Tirole. Tirole considers a three-level hierarchy with one principal, one supervisor, and one worker. The principal suffers from an adverse selection problem concerning the agent. Although the supervisor can help him partially overcome this asymmetry of information, the supervisor may withhold his information. Tirole studies the optimal contracts when the principal takes into account that the supervisor and the worker might play cooperatively, with covert transfers the game form he imposes. He shows in particular that the worker's incentive contract must be altered to make it coalition-proof.²

Note that the problem of coalition—inecentive compatibility arises as soon as there are more than two agents, as in the classical free-rider problem with public goods (see Green and Laffont, 1979).

In this article I point out and study a more general problem arising in organizations when incentive structures are not monitored by principals, as is necessarily the case as soon as the size of a firm increases sufficiently. Not only must the principal worry about coalitional behavior, with or without hidden transfers such as those analyzed by Tirole, but he must also consider the generalized notion of moral hazard that I call hidden gaming. It refers to the ability that some players may have to design and play games with other members of the hierarchy that are not observable by the principal. More precisely, the principal must worry that the incentive schemes he has set up do not create opportunities for some agents to organize hidden games by which they benefit from others. It is an extended notion of moral hazard, since the setting up of these hidden games can be viewed as an unobservable moral-hazard action.

This negative aspect of incentive schemes, well understood by union leaders, has not yet received any formal treatment. It is part of the influence costs stressed by Milgrom and Milgrom and Roberts. Also, it corresponds to the second sort of discretion envisioned by Crozier in The Bureaucratic Phenomenon. As he puts it: “We can envisage two sorts of discretion within an organization. The first comes from the uncertainty of the task itself, and the second from the rules that have been devised to make it more rational and more predictable.”

Drawing from the sociology literature, and in particular from Edwards, I have provided illustrations of the discretionary power of supervisors or foremen who often use it to their own benefit rather than to the benefit of the

² We do not consider the interesting issues arising from the need to induce cooperation between workers at the same time as using incentive schemes that create competition between them (see Crémèr, 1996; Aghion and Caillaud; Aoki).
organization. Sexual harassment, bribes, favors to relatives, etc., are part of a large class of transfers that affect the workings of organizations and that cannot be ignored in the design of incentive schemes (Laffont, 1988).

In his study of Japanese firms, Aoki (p. 111) stresses the problems raised by the important role of supervisor's soft information in evaluating workers:3 "one cannot deny that there is danger of unfair treatment of subordinates by a supervisor in the merit assessment."

In Section 2, we set up and motivate a simple model with a principal, a supervisor, and two workers. Workers affect production by their effort levels, but the principal can only observe total production. Without a supervisor, the principal can only tailor incentive schemes on total production. The supervisor's information enables him to construct more powerful incentive schemes based on individual performances. In Section 3, it is shown that with a benevolent supervisor the optimal incentive schemes for workers would use only individual performances (i.e., would be purely personalized).

In Section 4, hidden gaming is analyzed when the supervisor is not benevolent and his information is partially soft. More precisely, we introduce an index, $\delta$, of softness for the supervisor's information. This index admits several interpretations: $\delta$ can be viewed as the probability that the supervisor is not benevolent; or, assuming that all supervisors are not benevolent, it can also be the probability that the supervisor will be technically able to manipulate the information. If $\delta = 0$, we have the case of hard information; if $\delta = 1$, the pure case of soft information.

In the case of hard information ($\delta = 0$), we obtain the same result as if the supervisor were benevolent; the optimal schemes are purely personalized. However, as soon as $\delta$ becomes positive, the optimal mechanism involves some use of aggregate information about production (remember that aggregate information cannot be manipulated). For sufficiently large $\delta$ (larger than $\delta^* < \frac{1}{2}$), the optimal scheme relies on no personalized information. The main result is that hidden gaming makes profitable the use of information not valuable in the Blackwell sense (here, information about aggregate production).

In Section 5, the analysis is extended to a model with continuously variable effort levels. It is then possible to influence the hidden-gaming activity through its effect on effort. By making the supervisor's reward a function of total production, the supervisor's hidden-gaming activity can be controlled. However, unless the supervisor is infinitely risk-averse, some hidden-gaming activity remains for the optimal incentive structure, and the main conclusions presented in Section 4 hold.

3. This emphasis is partly due to the job rotation system, which makes the verifiable measurement of individual output difficult. As job rotation is an essential ingredient of Japanese firms' flexibility, we see that much research needs to be done on the interaction of incentive schemes and production processes.
2. THE MODEL

We consider a three-level hierarchy with a principal, a supervisor, and two workers. Each worker $i$, $i = 1, 2$, has the choice between two effort levels, $e^i \in \{0, 1\}$. At the zero-effort level he produces nothing. At the one-effort level he produces $x$ with probability $\pi$. His Von Neumann–Morgenstern (VNM) utility function is

$$U(s', e^i) = u(s') - \psi e^i,$$

where $s'$ is the monetary transfer he receives from the principal and $\psi$ is his disutility of effort. Finally, each worker has outside opportunities which give him an individual rationality (IR) level of utility, $W$.

Let us denote by $x^i$, with values in $\{0, x\}$, the random variable that represents worker $i$'s production level. We assume that $x^1$ and $x^2$ are independent random variables.

We also assume that there is a need for a supervisor to coordinate various activities within the organization. His VNM utility function is $v(t)$, $v' > 0$, $v'' < 0$, where $t$ is the monetary transfer he receives from the principal. Let $V$ be his individual rationality level of utility.

The principal is risk-neutral and his ex post utility level is

$$W = x^1 + x^2 - s^1 - s^2 - t.$$

We come now to the crucial part of the model which defines the information structures of the members in the organization. The supervisor observes individual production levels. Whereas his information about total production is "hard" (i.e., verifiable by the principal), so that the supervisor can be viewed as a technical tool enabling the principal to be informed about total production, the supervisor's information about individual production levels is partially soft (i.e., not verifiable by the principal or any third party).

We also assume that workers do not have hard information about their individual production levels. The following example is a motivation for our modeling. If the worker does not make any effort, the product is faulty and must be rejected. Even if he does make an effort, the product is faulty with probability $1 - \pi$. However, this faulty nature of the product is ascertained by a testing process that is in the hands of the supervisor. In this procedure the supervisor cannot misrepresent aggregate results, so that total production is hard information; however, he can switch the results concerning

---

4. Demski and Sappington provide a similar analysis in the case of the government-regulator-firm hierarchy.

5. In Section 5, we extend the model to a continuous-effort variable.

6. In this way we formally exclude any technical need for cooperation between workers.
different workers. Edwards gives clear evidence of the relevance of this setup when he discusses the role of foremen and supervisors at the Digitex company.

The supervisors' immediate role in directing production gives them considerable power, of course, yet their full power springs from other sources as well . . . foremen maintain a certain degree of control because they must approve any "benefits" the workers receive. . . For piece-rate workers, who are eligible for raises, the supervisors' decisions on rejects—what to count as faulty output and whether to penalize the workers for it—weigh heavily in bonus calculations.

3. FIRST BEST AND BENEVOLENT SUPERVISOR

Given the technology, it is essential for the principal to provide incentives for workers' effort, whatever his own information structure. If \( x \) is large enough, inducing effort is indeed optimal.

Suppose, as a benchmark, that the principal has complete information about outputs and effort levels. In that event, the optimal rewards are

\[
\begin{align*}
    s^1 &= s^2 = s^* = u^{-1} (U + \psi) = \varphi (U + \psi), \\
    t &= v^{-1} (V) = \xi (V).
\end{align*}
\]

Both workers make efforts and are compensated for it. The workers and the supervisor are set at their IR levels of utility.

Assume instead that the principal observes nothing, but that his supervisor is benevolent\(^7\) and does not need to be motivated to provide truthfully his soft information. From the tournaments literature (Lazear and Rosen, Mookherjee, Green and Stockey), we know that the independence of the random variables \( x^1 \) and \( x^2 \) implies the optimality of (identical) personalized incentive schemes\(^8\) [i.e., of workers' rewards \( s(x^i) \)].

Since the principal must elicit the level-one effort for his organization to be valuable, and since he will want to saturate workers' IR constraints, the optimal mechanism for each worker is simply defined from the IR and incentive-compatibility (IC) constraints.

Let us denote for a given worker \( i \) state 0 (state 1) as the state \( x^i = 0 (x^i = \bar{x}) \) and let

\[
\begin{align*}
    s_0 &= s (0), & u_0 &= u (s_0), \\
    s_1 &= s (\bar{x}), & u_1 &= u (s_1).
\end{align*}
\]

\(^7\) Alternatively, we could assume that his information about individual production levels is hard as well.

\(^8\) We will prove this result for our model in Section 4.
Then, the IC constraint is
\[ \pi u_1 + (1 - \pi) u_0 - \psi \geq u_0. \] (3)

The IR constraint is
\[ \pi u_1 + (1 - \pi) u_0 - \psi \geq U. \] (4)

The benevolent supervisor is set at his IR level. The principal maximizes his expected utility under (3) and (4):
\[ 2 [\pi \bar{x} - \pi \varphi (u_1) - (1 - \pi) \varphi (u_0)] - \xi (\bar{Y}). \] (5)

We get immediately
\[ u_0 = U, \quad u_1 = U + \psi/\pi. \] (6)

For further reference we gather the following results.

**Proposition 1.** With a benevolent supervisor, the optimal incentive schemes of the workers are personalized and are characterized by
\[ s_0 = \varphi (U), \] (7)
if the worker's production is 0,
\[ s_1 = \varphi (U + \psi/\pi), \] (8)
if the worker's production is \( \bar{x} \).

**Remark.** Not observing effort levels costs the principal
\[ 2 [\pi s_1 + (1 - \pi) s_0 - s^*] > 0, \]
from the strict convexity of \( \varphi \) and Jensen's inequality.

If he relies only on hard information, the principal can design nonpersonalized incentive schemes for workers. Three levels of total production are possible, \( X \in \{0, \bar{x}, 2\bar{x}\} \). The total compensation provided by the principal is now a function of \( X \) (we still denote it by \( s \)) and, by symmetry,\( ^{8} \) it is shared equally between the workers. Let
Worker $i$’s objective function is now

$$u[\frac{1}{2}x_i + x_j] - \psi e^i, \quad i = 1, 2, \quad j \neq i.$$  

We assume that effort levels result from a Nash equilibrium, taking incentive schemes as given. Each worker’s IR and IC constraints are, respectively,

$$\begin{align*}
(1 - \pi)^2 u_0 + 2\pi(1 - \pi)u_1 + \pi^2 u_2 - \psi &\geq U, \\
(1 - \pi)^2 u_1 + 2\pi(1 - \pi)u_2 + \pi^2 u_2 - \psi &\geq \pi u_1 + (1 - \pi)u_0.
\end{align*}$$

The principal’s optimization program can then be written as

$$\begin{align*}
\max \{ 2\pi x - 2(1 - \pi)^2 \varphi(u_0) - 4\pi(1 - \pi)\varphi(u_1) - 2\pi^2 \varphi(u_2) - \xi(v) \},
\end{align*}$$

such that

$$\begin{align*}
\pi^2 u_2 + 2\pi(1 - \pi)u_1 + (1 - \pi)^2 u_0 - \psi &\geq U (\lambda), \\
\pi^2 u_2 + \pi(1 - 2\pi)u_1 - \pi(1 - \pi)u_0 - \psi &\geq 0 (\mu), \\
v &\geq \underline{V}.
\end{align*}$$

The optimal solution is summarized in the following proposition.

**Proposition 2.** When the principal relies only on hard information, the optimal incentive schemes of the workers are nonpersonalized and are characterized by

$$\begin{align*}
\pi u \left( \frac{s_1}{2} \right) + (1 - \pi) u \left( \frac{s_0}{2} \right) &= U, \\
\pi u \left( \frac{s_2}{2} \right) + (1 - \pi) u \left( \frac{s_1}{2} \right) &= U + \psi/\pi,
\end{align*}$$

9. Here and below, symmetry is a result and not an assumption. This is easily seen by observing that, because of the stochastic independence of the random shocks, the optimization program of the principal can always be decomposed into one program for each worker with the same optimal solution.
Proof. From our assumptions about \( u(\cdot) \) and \( v(\cdot) \) we are maximizing in \((u_0, u_1, u_2, v)\) a strictly concave utility function with linear constraints. The solution is characterized by the first-order conditions. Note first that \( v = V \). Then we have

\[
-2(1 - \pi)^2 \phi'(u_0) + (1 - \pi)^2 \lambda - \pi(1 - \pi)\mu = 0, \quad (12)
\]

\[
-4\pi(1 - \pi)\phi'(u_1) + 2\pi(1 - \pi)\lambda + \pi(1 - 2\pi)\mu = 0, \quad (13)
\]

\[
-2\pi^2 \phi'(u_2) + \pi^2 \lambda + \pi^2 \mu = 0. \quad (14)
\]

It is easily checked that both IR and IC constraints must be saturated. From these two equations we obtain

\[
\pi u_1 + (1 - \pi) u_0 = U, \quad (15)
\]

\[
\pi u_2 + (1 - \pi) u_1 = U + \psi/\pi. \quad (16)
\]

Then, combining (12)-(14) we get

\[
\phi'(u_1) = \phi'(u_2) + \phi'(u_0)/2. \quad (17)
\]

Q.E.D.

The reward function is computed to induce effort. Comparing (16) with (15), we observe that the worker gets an expected utility of \( U + \psi/\pi \) when exerting effort versus \( U \) when not. That is similar to (7) and (8), except that it is subject to the randomness of the other worker's output (for his equilibrium effort level). Finally, the minimization of expected reward under IC and IR constraints is achieved by the simple formula (17).

4. SUPERVISOR'S HIDDEN GAMING

We now consider the case where the supervisor's information concerning individual performance is partially soft, and the supervisor is not benevolent. Here, \( \delta \) is an index of softness such that, if \( \delta = 0 \), the information about individual production levels is hard, and, if \( \delta = 1 \), it is soft (i.e., nonverifiable).

Note first that, in our particular model, the supervisor cannot benefit from forming a coalition with both workers. Even if he uses personalized
incentive schemes, the principal can determine his total wage bill, $S$, on the basis of his hard information, $X$. If $X = 2\bar{x}; \bar{z}; 0$, then it can be inferred $S = 2s_1; s_1 + s_0; 2s_0$. Therefore, by misrepresenting his soft information (namely, individual performances), the supervisor cannot improve the wage bill and therefore cannot benefit from it.

However, the supervisor can engage in other forms of hidden behavior. First, he can form a coalition with one worker at the expense of the other worker. Also, he can extract benefits from each worker by threatening to favor the other worker. The extent to which these strategies work depends crucially on the commitment abilities of both the principal and the supervisor.

At this stage of development of the theory of organizations based on informational asymmetries, we must admit the necessity of some ad hoc assumptions on commitment abilities. A deeper theory would recast the analysis in its historic dimension. The emergence of various forms of reputation making credible certain types of commitment would then be a result of this dynamic framework. However, we believe that the results would then be very dependent on prior beliefs and, without a sociological theory about these beliefs, the ad hoc basis of the commitment assumptions would simply be replaced by the ad hoc basis of prior beliefs. Our approach is to make the commitment assumptions that appear relevant in view of our empirical knowledge of the institutions we study. A systematic study of the relation between these commitment assumptions and the functioning of the institutions seems to us the most fruitful research agenda.

At one extreme, if the principal can commit to a scheme that leaves no surplus to workers, any attempt by the supervisor to extract benefits from them will push the workers' welfare below their IR level and out of the organization, and therefore will be self-destructive. In other words, extraction of benefits could not be a subgame perfect equilibrium. We will assume that the principal cannot commit to incentive schemes for workers and that, following the supervisor's choice of collusive behavior or extortion, he will renegotiate contracts with workers that keep them in the organization. The supervisor will rationally anticipate this renegotiation. As a first step, we assume that no attempt is made to monitor the supervisor's activity by making his transfer a function of observed variables or messages. We get the following timing:

<table>
<thead>
<tr>
<th>Contract with workers and supervisor</th>
<th>Renegotiation with workers</th>
<th>Production levels take place</th>
</tr>
</thead>
<tbody>
<tr>
<td>hidden gaming</td>
<td></td>
<td>if necessary</td>
</tr>
</tbody>
</table>

10. A simple justification for this assumption arises when the principal must sink large resources into his relationship with the supervisor. Then, the threat of dismissing the supervisor if workers quit may not be credible.
Suppose that we have the most general incentive scheme for each worker, which relies on both aggregate information and the individual performances reported by the supervisor, \( \hat{x}_1, \hat{x}_2 \) (where a caret refers to announced performances).

If

\[
\begin{align*}
X &= 0, \quad s^1 = s^2 = s_0/2, \quad u_0 = u(s_0/2); \\
X &= \hat{x}, \quad \hat{x}_1 = \hat{x}, \quad \hat{x}_2 = 0, \quad s^1 = s_{11}, \quad s^2 = s_{10}; \\
& \quad u_{11} = u(s_{11}), \quad u_{10} = u(s_{10}); \\
X &= \hat{x}, \quad \hat{x}_1 = 0, \quad \hat{x}_2 = \hat{x}, \quad s^1 = s_{10}, \quad s^2 = s_{11}; \\
& \quad u_{10} = u(s_{10}), \quad u_{11} = u(s_{11}); \\
X &= 2\hat{x}, \quad s^1 = s^2 = s_2/2, \quad u_2 = u(s_2/2).
\end{align*}
\]

First, consider the coalition of the supervisor and one worker, say, worker 1. The coalition can only gain in state 1 of aggregate information when \( x_1 = 0 \) and \( x_2 = \hat{x} \). Then, the supervisor can claim that production has been achieved by worker 1 instead of worker 2. The expected gain for the coalition is

\[
\delta \pi (1 - \pi)(u_{11} - u_{10}).
\]

which is the expected utility gain of worker 1 from this distortion, where \( \delta \) is an index of softness of the supervisor's information. This gain is split between the supervisor and worker 1 in a way that does not have to be specified because it does not affect our subsequent analysis.

As worker 2 is not approached by the supervisor, he anticipates that worker 1 and the supervisor collude and, therefore, that he will lose, in expected terms, \( \delta \pi (1 - \pi)(u_{11} - u_{10}) \). This quantity must be subtracted from his expected utility in order to write the IR and IC constraints:

\[
(1 - \pi)^2 u_0 + \pi (1 - \pi)(u_{11} + u_{10}) + \pi^2 u_2 - \psi \\
- \delta \pi (1 - \pi)(u_{11} - u_{10}) \geq U, \\
\]

\[
(1 - \pi)^2 u_0 + \pi (1 - \pi)(u_{11} + u_{10}) + \pi^2 u_2 - \psi \\
- \delta \pi (1 - \pi)(u_{11} - u_{10}) \geq (1 - \pi)u_0 + \pi u_{10}.
\]

Equation (20) incorporates the fact that the worker cannot be cheated if he does not exert any effort.

The principal, without knowledge of who is colluding with the supervisor, must renegotiate the workers' contract so that (19) and (20) are satisfied for both workers.\(^{11}\)

11. Worker 1's true incentive and individual rationality constraints are then automatically satisfied.
Instead of colluding with one worker, the supervisor can face the two workers with the following prisoner-dilemma game. Each worker is told that if he does not provide a favor of $b$ to the supervisor, the other worker will be credited the benefit of production $\mathit{x}$ even if it comes from the first worker.

Each worker then has two strategies: giving in ($G$) or resisting ($R$). To give in means providing benefits (say, $b^i$ for worker $i$), and resisting means not providing the benefits. If both resist, the supervisor has no particular reason to discriminate, and we assume that he reports truthfully. The payoffs to workers are summarized in the display below.\(^{12}\)

<table>
<thead>
<tr>
<th>Worker 1</th>
<th>Worker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>$U - b^1$</td>
<td>$U - \delta\pi(1 - \pi)(u_{i1} - u_{i0})$</td>
</tr>
<tr>
<td>$U - b^2$</td>
<td>$U + \delta\pi(1 - \pi)(u_{i1} - u_{i0}) - b^2$</td>
</tr>
<tr>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>$U + \delta\pi(1 - \pi)(u_{i1} - u_{i0}) - b^1$</td>
<td>$U$</td>
</tr>
<tr>
<td>$U - \delta\pi(1 - \pi)(u_{i1} - u_{i0})$</td>
<td>$U$</td>
</tr>
</tbody>
</table>

Here $\bar{U}$ denotes the expected utility that workers retain from the undistorted incentive schemes.

It is a dominant strategy for each worker to give in as long as

$$b^i \leq \delta\pi(1 - \pi)(u_{i1} - u_{i0}), \quad i = 1, 2.$$  

Then, the supervisor gains $b^1 + b^2$. If he is not motivated otherwise, the supervisor will choose a maximal extortion with a benefit of $2\delta\pi(1 - \pi)(u_{i1} - u_{i0})$.

Note that the constraints in the renegotiation are still (19) and (20), with the only difference that now these constraints are relevant for the two workers.

The next proposition characterizes the optimal (renegotiation-proof) incentive schemes when one of these forms of hidden gaming takes place.

**Proposition 3.** When the supervisor is not benevolent and the softness index of his information is $\delta$, the optimal incentive scheme for workers is such that

\(i\) \quad (1 - \pi)u(s_0/2) + \pi u(s_{i0}) = U;

\(^{12}\) Benefits do not have to be pecuniary, and the gain to the supervisor does not have to equate the loss to the worker as we assume here for definiteness. Also, the issue arises of how the supervisor can credibly commit to the payoffs defining the hidden game. We must appeal to reputation since other forms of commitment are not available given the forbidden nature of these transactions. The same problem arises for coalitional behavior of the ability to commit to (illegal) covert transfers.
\( \delta(1 - \pi)u(s_{10}) + (1 - \delta)\pi u(s_{11}) + \pi u(s_{2}/2) = U + \psi/\pi; \)
\( \frac{1}{u'(s_{11})} + \frac{1}{u'(s_{10})} = \frac{1}{u'(s_{2}/2)} + \frac{1}{u'(s_{2}/2)}; \)

(iii) for \( \delta = 0 \), the optimal scheme is purely personalized;

(iv) for \( \delta > 0 \), optimal mechanisms use aggregate information;

(v) there exists \( \delta^* < \delta \), such that the optimal scheme is nonpersonalized \( u_{11} = u_{10} \) iff \( \delta > \delta^* \).

Proof. The principal’s optimization program can be rewritten
\[
\max_{\{u_0, u_{10}, u_{11}, u_2\}} \left\{ 2\pi \xi - 2\pi^2 \phi(u_2) - 2\pi(1 - \pi)[\phi(u_{10}) + \phi(u_{11})] \right. \\
- 2(1 - \pi)^2 \psi(u_0) - \xi(V) \right. \],
\]
such that
\[
(1 - \pi)^2 u_0 + (1 + \delta)\pi(1 - \pi)u_{10} + (1 - \delta)\pi(1 - \pi)u_{11} + \pi^2 u_2 \geq U + \psi \quad (\lambda), \tag{21}
\]
\[
(1 - \pi)u_0 + \pi u_{10} = U \quad (\mu), \tag{22}
\]
\[
u_{11} - u_{10} \geq 0 \quad (\nu). \tag{23}
\]

Equation (22) is a combination of the original constraints, which are easily seen to be tight. Also, we incorporate the fact that the supervisor’s transfer has been committed to the IR level \( \xi(V) \).

The first-order conditions of this concave problem are
\[-2\pi^2 \phi'(u_2) + \lambda \pi^2 = 0, \tag{24}
\]
\[-2\pi(1 - \pi)\phi'(u_{10}) + \lambda (1 + \delta)\pi(1 - \pi) + \mu \pi - \nu = 0, \tag{25}
\]
\[-2\pi(1 - \pi)\phi'(u_{11}) + \lambda (1 - \delta)\pi(1 - \pi) + \nu = 0, \tag{26}
\]
\[-2(1 - \pi)^2\phi'(u_0) + \lambda (1 - \pi)^2 + \mu (1 - \pi) = 0. \tag{27}
\]

Substituting (22) into (21) leads to
\[
\delta(1 - \pi)u_{10} + (1 - \delta)(1 - \pi)u_{11} + \pi u_{2} = U + \psi/\pi. \tag{28}
\]
Adding (25) to (26) and using (24) and (27) give

\[ \varphi'(u_{10}) + \varphi'(u_{11}) = \varphi'(u_0) + \varphi'(u_2). \]  

(29)

Two cases may occur according to \( v \): case a, \( v > 0 \) [then \( u_{11} = u_{10} \) (i.e. the scheme is nonpersonalized)]; or case b, \( v = 0 \),

\[ \varphi'(u_{11}) = (1 - \delta)\varphi'(u_2), \]  

(30)

\[ \varphi'(u_{10}) = \varphi'(u_0) + \delta \varphi'(u_2). \]  

(31)

For this case to occur we need \( u_{11} > u_{10} \).

For \( \delta = 0 \), \( u_0 = u_{10} \), \( u_{11} = u_{12} \). We prove here that, with a benevolent supervisor, or when his personalized information is hard, optimal schemes are personalized (see Proposition 2).

For \( \delta > 0 \), but small (by continuity we are in case b), we use both aggregate and individualized information.

For \( \delta > \frac{1}{2} \), case b is impossible [see (30) and (31)], and the principal switches to nonpersonalized incentive schemes. We prove now that there exists a value \( \delta^* \in [0, \frac{1}{2}] \) such that optimal contracts are nonpersonalized iff \( \delta \geq \delta^* \).

From the envelope theorem, if \( W(\delta) \) is the expected value of the principal’s program as a function of \( \delta \), we have

\[ \frac{dW(\delta)}{d\delta} = -\lambda \pi (1 - \pi) (u_{11} - u_{10}). \]

As long as schemes are personalized, \( u_{11} > u_{10} \), the value strictly decreases with \( \delta \) (since \( \lambda > 0 \)). The value of a nonpersonalized incentive scheme is independent of \( \delta \) and we know that, for \( \delta \) large enough, such a nonpersonalized scheme is optimal. Therefore there exists \( \delta^* \) such that a nonpersonalized scheme is optimal iff \( \delta \geq \delta^* \).

Q.E.D.

Proposition 3 is illustrated by Figure 1, where \( W \) is the principal’s expected utility for the optimal nonpersonalized incentive scheme.

In Proposition 3, (i) and (ii) structure the contract, such that effort is achieved by both workers; (iii), the analog of (ii) in Proposition 2, expresses cost minimization by the principal; and (iv) proves the statement made earlier that with independent production processes aggregate information is useless when there is no hidden gaming. On the other hand, (iv) proves that, as soon as some hidden gaming occurs, aggregate information is useful. Here we point out a new role for aggregate information even when production processes are independent. Then, (vi) shows that, if hidden gaming is too severe, the principal is better off giving up the supervisor’s soft information. Finally, let us note that, for a whole range of values for \( \delta \), some extortion
Figure 1.

activity subsists at the optimal mechanism since it is somewhat personalized.

Is it possible to do better by motivating the supervisor? Any attempt to make the supervisor’s transfer a function of total production has no effect as long as (as we have postulated) the supervisor anticipates renegotiation, and this renegotiation is indeed always valuable for the principal. Any attempt to make his transfer a function of his announcement fails also because he can make only two announcements in the critical state \( x_1 = x \) or \( x_1 = 0 \), and if he is not motivated to make a variable statement, then his soft information becomes valueless.

On the other hand, he could be ex ante penalized from the anticipated level of extortion he will realize at the equilibrium. In this way the principal would gain back \( \delta(1 - \pi)\pi(u_{11} - u_{10}) \) [or \( 2\delta(1 - \pi)\pi(u_{11} - u_{10}) \) in the hidden-gaming case].

This would slightly affect Proposition 3 by boosting somewhat the personalization of optimal contracts. Parts (i)–(iv) still hold, (v) will still be generically true, and (vi) may be false. In particular, in view of this additional term, nonpersonalized schemes may never be optimal for \( \delta < 1 \). This last feature, surprisingly, happens more often with hidden-gaming than coalitional behavior.

In all cases, it remains that hidden gaming yields a distortion of optimal incentive schemes which incorporate the use of aggregate information. Note that this is different from Holmstrom’s point that additional information correlated with outputs should be used in designing mechanisms in moral hazard contexts. Without hidden gaming this would not be the case in our model, because this additional information would not be valuable in the Blackwell sense, and personalized schemes without any use of aggregate production information would be optimal.
The analysis so far has been greatly simplified by the \( \{0, 1\} \) nature of effort. We examine in the next section the difficulties raised by relaxing this hypothesis.

5. CONTINUOUS EFFORT LEVEL

We extend the model by assuming that the probability \( \pi \) is now an increasing function of the effort level, which is a continuous variable in \( \mathbb{R}_+ \). There exists \( \varepsilon > 0 \), such that \( \pi(e) \) in \( [e, 1 - \varepsilon] \) for any \( e \) in \( \mathbb{R}_+ \). We further assume that \( \pi''(e) < 0, e \in \mathbb{R}_+ \) to validate the first-order approach we use in this article (see Appendix A for more details).

The IR and IC constraints of a worker who is submitted to extortion,\(^\text{13}\)
\[
\delta \pi(e)(1 - \pi(e))(u_{11} - u_{10}) \geq \psi(e) \geq U, \tag{32}
\]

\[
\pi'(e)(1 - \pi(e))u_{10} + \pi(e)u_{11} \geq \psi(e) = U. \tag{33}
\]

For our formulation to be meaningful, we also impose the constraint
\[
u_{11} - u_{10} \geq 0. \tag{34}
\]

Let us still assume that the supervisor is paid a constant wage \( \xi(V) \). Then, the principal maximizes his expected utility,
\[
2\pi(e)\xi - 2(1 - \pi(e))^2\varphi(u_0) - 2\pi(e)(1 - \pi(e))(\varphi(u_{10}) + \varphi(u_{11})) - 2\pi(e)^2\varphi(u_2) - \xi(V), \tag{35}
\]
under the constraints (32)–(34).\(^\text{14}\) We obtain a result similar to Proposition 3.

**Proposition 4.** When the supervisor is paid a constant wage and is not benevolent, the optimal incentive contracts for workers are such that

(i) if \( \delta = 0 \), the optimal contracts are purely personalized;

(ii) if \( \delta > 0 \), the optimal contracts use aggregate data;

\(^{13}\) In this section, we consider only the case of hidden gaming. The analysis of coalitional behavior is complicated by the fact that workers are not in a symmetric situation and do not choose the same effort levels.

\(^{14}\) The second-order condition is satisfied if \( u_{11} \geq u_{10} \), \( u_{11} \geq u_{10} \), and \( u_{11} \geq u_{10} \). We impose \( u_{11} \geq u_{10} \), and we verify that the other constraints are satisfied at the (unconstrained) optimum.
Proof. See Appendix B.

In this new model, hidden gaming affects not only the principal's expected utility, but also the equilibrium effort level and, therefore, the probability distribution of production. It then becomes possible to affect the behavior of the supervisor by making his wage a function of the production level.

The supervisor will now select his level of extortion \( b \) in the interval \([0, \delta(u_{11} - u_{10})]\), knowing that his extortion affects effort and therefore his own rewards. Let this reward be

\[
    t_0 \text{ if } X = 0; \quad t_1 \text{ if } X = \tilde{x}; \quad t_2 \text{ if } X = 2\tilde{x}.
\]

For a given level of extortion \( b \), workers choose their effort level according to

\[
    \pi'(e)(1 - \pi(e))(u_{11} - u_0 - b) + \pi(e)(u_2 - u_{10}) - \psi'(e) = 0, \quad (36)
\]

which implicitly defines a function \( e^*(b/\tilde{u}) \) with \( \tilde{u} = (u_0, u_{10}, u_{11}, u_2) \). The supervisor anticipates this reaction taking the workers' contract as fixed but assuming that the IR constraint of the worker always will be satisfied by renegotiation. Therefore, he chooses his level of extortion by solving

\[
    \max_{\{t_0, t_1, t_2\}} \{2b + (1 - \pi(e))^2\nu(t_0) + 2\pi(e)(1 - \pi(e))\nu(t_1) + \pi(e)^2\nu(t_2)\},
\]

such that

\[
    0 \leq b \leq \delta(u_{11} - u_{10}) \text{ and } (36).
\]

Two main cases may occur. Either the optimal choice of \( b \) saturates the constraint or it does not. Under saturation, the transfers \((t_0, t_1, t_2)\) do not succeed in modifying the supervisor's extortion behavior. Then the supervisor is paid a constant wage (because this is the cheapest way in view of his risk aversion to provide his IR level of utility). The workers' optimal contracts are then characterized by Proposition 4.

We will concentrate on the other case when the optimal \( b \) does not achieve its boundary value. We will assume that the optimal \( b \) is characterized by the first-order condition of the supervisor's program:
\[ 0 = 2 + \frac{\partial e^*}{\partial b} \left( \frac{b}{u} \right) \pi'(e)[-2(1 - \pi(e))v(t_0) + 2(1 - 2\pi(e))v(t_1) + 2\pi(e)v(t_2)]. \]  

(37)

The renegotiation-proof contracts are then obtained by the principal from the program

\[
\text{Max} \quad \{2\pi(e)x - (1 - \pi(e))^2(t_0 + 2\varphi(u_0))
\]

\[
- 2\pi(e)(1 - \pi(e))(t_1 + \varphi(u_{11}) + \varphi(u_{10}))
\]

\[
- \pi(e)^2(t_2 + 2\varphi(u_2)),
\]

such that

\[
(1 - \pi(e))^2u_0 + \pi(e)(1 - \pi(e))[u_{10} + u_{11} - b] + \pi(e)^2u_2 - \psi(e) = U,
\]

\[
\pi'(e)[(1 - \pi(e))(u_{11} - u_0 - b) + \pi(e)(u_2 - u_{10})] - \psi'(e) = 0,
\]

\[
2b + (1 - \pi(e))^2v(t_0) + 2\pi(e)(1 - \pi(e))v(t_1) + \pi(e)^2v(t_2) = V.
\]

Contrary to the case of a binding constraint, we can now show that the workers' optimal contract is affected by the principal's control of the supervisor. The supervisor's contract can be viewed as a classic principal-agent contract in which the supervisor maximizes, with respect to the workers' equilibrium effort level, the difference between the expected utility of his transfer and the cost of effort [which, for him, is \(-2b(e)\)] minus the extortion he will be able to get as a function of the workers' effort levels.

The solution to the principal's optimization program can be fairly complex and unintuitive, because the dependence of the effort level with respect to extortion can be increasing or decreasing. We will concentrate below on the normal case, where effort decreases with extortion. Moreover, we will assume that we are not at a corner solution, \(e^* = 0\), and that the first-order approach is valid. We then obtain the following.
Proposition 5. Under the above regularity assumptions,

(i) the incentive contracts of workers are not purely personalized and \( u_{10} < u_0 \),

(ii) the supervisor's reward increases with production.

Proof. See Appendix C.

With Propositions 4 and 5, we have extended our former result that hidden gaming leads to the use of aggregate production information in the workers' incentive contracts, despite the independence of their production processes. In the case where the amount of extortion is not limited by \( u_{11} - u_{10} \), the spread \( u_{11} - u_{10} \) is enlarged to the point of making the reward to the worker who does not produce when the other is producing worse than when they both do not produce (\( u_{10} < u_0 \)).

Finally, it is clear that unless the risk aversion of the supervisor is infinite, there is no reason for the level of extortion at the optimum to be 0.

6. CONCLUSION

The extension of the principal agent methodology to include a third level, here represented by a supervisor, raises a host of difficult questions. A number of assumptions have been necessary to carry out our program, and our model does not pretend to cover all the aspects of three-level hierarchies.

We assume that workers themselves do not form a coalition in reaction to the supervisor's behavior. Note that if workers make covert transfers between themselves, only nonpersonalized incentive schemes are possible. However, this behavior is made difficult when rewards and transfers are unobservable. Also, we assume that workers are unable to complain directly to the principal, and have no other appeals (such as a grievance proceeding). The way management attempts to control supervisors is an interesting topic in itself. We assume that the size of the firm is such that this type of control is difficult.

Our major conclusions are (a) extortion activity by the supervisor may appear in a hierarchy and raises the issue of improved private-ordering public intervention; (b) hidden gaming may make nonpersonalized incentive schemes optimal; and (c) hidden gaming justifies the use of aggregate production information in individual incentive schemes even when it is not Blackwell-relevant.
APPENDIX A: VALIDITY OF THE FIRST-ORDER APPROACH

From Mirrlees and Rogerson we know that the first-order approach is valid in the principal–one-agent model, when an increase of $e$ creates a first-order stochastic dominance improvement in the distribution of output [if $F(x,e)$ is the cumulative distribution function of $x$ for an effort $e$, this requires $F_e(x,e) \leq 0$ for any $x,e$, and strict inequality for a set of positive measure]; and if $F$ is convex in $e$, $F_{ee} \geq 0$ with the monotone likelihood ratio hypothesis: $F_e(x,e)/F(x,e)$ nondecreasing in $x$. At least it is true in the sense that if we are not at the corner solution, $e = 0$, the first-order approach is valid.

Our distribution indeed satisfies all these properties.

The extension to multiple agents with independent production processes is trivial, since the optimal mechanisms are based on individual performances just as in the one-agent case. However, in our case, hidden gaming requires the use of aggregate information and the extension is not obvious.

We must now distinguish four states for a given agent $i$:

\[
\begin{align*}
  w_1 &= \text{state 1, } X = 0; \\
  w_2 &= \text{state 2, } X = \tilde{x} \text{ with } x' = 0; \\
  w_3 &= \text{state 3, } X = \tilde{x} \text{ with } x' = \tilde{x}; \\
  w_4 &= \text{state 4, } X = 2\tilde{x}.
\end{align*}
\]

The states $w_i$ are cardinalized in such a way that $w_1 < w_2 < w_3 < w_4$, where $\pi$ is the probability that the other worker produces $\tilde{x}$:

\[
\begin{align*}
  \Pr(w_1/e) &= (1 - \pi(e))(1 - \pi), \\
  \Pr(w_2/e) &= (1 - \pi(e))\pi, \\
  \Pr(w_3/e) &= \pi(e)(1 - \pi), \\
  \Pr(w_4/e) &= \pi(e)\pi;
\end{align*}
\]

\[
\begin{align*}
  F(w_1/e) &= (1 - \pi(e))(1 - \pi) \rightarrow F_e(w_1/e) = -\pi'(e)(1 - \pi) < 0, \\
  F(w_2/e) &= (1 - \pi(e)) \rightarrow F_e(w_2/e) = -\pi'(e) < 0, \\
  F(w_3/e) &= (1 - \pi(e))\pi + (1 - \pi) \rightarrow F_e(w_3/e) = -\pi'(e)\pi < 0, \\
  F(w_4/e) &= 1 \rightarrow F_e(w_4/e) = 0 \leq 0;
\end{align*}
\]

\[
\begin{align*}
  F_{ee}(w_1/e) &= -\pi''(e)(1 - \pi) \geq 0, \quad \frac{F_e(w_1/e)}{F(w_1/e)} = -\frac{\pi'(e)}{1 - \pi(e)}, \\
  F_{ee}(w_2/e) &= -\pi''(e) \geq 0, \quad \frac{F_e(w_2/e)}{F(w_2/e)} = -\frac{\pi'(e)}{1 - \pi(e)}.
\end{align*}
\]
From the point of view of an agent, the problem is identical (despite hidden gaming) to the usual case if we make the change of variables $\hat{u}_{10} = (1 + \delta)u_{10}$, $\hat{u}_{11} = (1 - \delta)u_{11}$. If we assume that the optimal solution is not the corner solution $e = 0$, and $\delta$ is small enough, then we show in Appendix B that the multiplier of the worker’s incentive constraint is positive and, therefore, that the reward to a worker is increasing in $w$ in view of the monotone likelihood-ratio hypothesis. Then, it is easy to check that the worker’s problem is concave in effort and, therefore, the first-order approach is validated. For larger $\delta$, we did not prove the validity of the first-order approach.

APPENDIX B: PROOF OF PROPOSITION 4

The first-order conditions reduce to (32), (33), and

\[ 2\psi'(u_0) = \lambda - \mu \frac{\pi'(e)}{1 - \pi(e)}, \]  
\[ 2\psi'(u_{10}) = (1 + \delta) \left[ \lambda - \mu \frac{\pi'(e)}{1 - \pi(e)} \right] - \frac{\nu}{2\pi(e)(1 - \pi(e))} + \mu \delta \frac{\pi'(e)}{\pi(e)(1 - \pi(e))}, \]  
\[ 2\psi'(u_{11}) = (1 - \delta) \left[ \lambda + \mu \frac{\pi'(e)}{\pi(e)} \right] + \frac{\nu}{2\pi(e)(1 - \pi(e))} - \mu \delta \frac{\pi'(e)}{(1 - \pi(e))\pi(e)}, \]  
\[ 2\psi'(u_2) = \lambda + \mu \frac{\pi'(e)}{\pi(e)}, \]  
\[ \nu(u_{11} - u_{10}) = 0, \quad \nu \geq 0, \quad \lambda \geq 0. \]

For $\delta$ close to 0, we can show (following the reasoning of Holmstrom, 1979) that $\mu > 0$. 
Then for $\delta = 0$, we can conclude that the scheme is purely personalized. Indeed, $\delta = 0$ implies $u_{10} < u_{11}$, which implies $\nu = 0$, which implies $u_{10} = u_0$ and $u_2 = u_{11}$.

For $\delta$ close to 0, if $u_{11} \neq u_{10}$, then $\nu = 0$. Thus, from (B1) and (B2), $u_0 \neq u_{10}$. If $u_{11} = u_{10}$, then aggregate information is not used if $u_0 = u_{11} = u_{10} = u_2$, which leads to a contradiction in view of $\nu > 0$.

If $\delta$ is close to 1 and $u_{11} > u_{10}$, then (B2) and (B3) imply $u_{11} < u_{10}$, a contradiction. Therefore, $u_{11} = u_{10}$ for $\delta$ close to 1.

From the envelope theorem

$$
\frac{dW}{d\delta} = -\lambda \pi(e)(1 - \pi(e))(u_{11} - u_{10}) - \mu \nu'(e)(1 - \pi(e))(u_{11} - u_{10}) < 0.
$$

Because for $\delta = 0$ the system is personalized, for $\delta$ close to 1 it is nonpersonalized, $dW/d\delta < 0$, and because the expected value of the nonpersonalized system is independent of $\delta$, there exists $\delta^* < 1$, such that the system is nonpersonalized iff $\delta \geq \delta^*$.

APPENDIX C: PROOF OF PROPOSITION 5

Let $\lambda^w$, $\lambda^s$, $\mu^w$, and $\mu^s$ be the multipliers of the IR and IC constraints of the workers (w) and the supervisor (s). The first-order conditions are

\begin{align*}
(u_0) &\quad -2(1 - \pi(e))\varphi'(u_0) + \lambda^w(1 - \pi(e))^2 - \mu^w\pi'(e)(1 - \pi(e)) = 0, \\
(u_{11}) &\quad -2\pi(e)(1 - \pi(e))\varphi'(u_{11}) + \lambda^w\pi(e)(1 - \pi(e)) \\
&\quad + \mu^w\pi'(e)(1 - \pi(e)) = 0, \\
(u_{10}) &\quad -2\pi(e)(1 - \pi(e))\varphi'(u_{10}) + \lambda^w\pi(e)(1 - \pi(e)) \\
&\quad - \mu^w\pi'(e)\pi(e) - 2\mu^s\pi'(e)/(1 - \pi(e))^2 = 0, \\
(u_2) &\quad -2\pi(e)^2\varphi'(u_2) + \lambda^w\pi(e)^2 + \mu^w\pi'(e)\pi(e) \\
&\quad + 2\mu^s\pi'(e)/(1 - \pi(e))^2 = 0, \\
(t_0) &\quad -(1 - \pi(e))^2 + [(1 - \pi(e))^2\lambda^s - \mu^s \cdot 2\pi'(e)(1 - \pi(e))]\nu'(t_0) = 0, \\
(t_1) &\quad -2\pi(e)(1 - \pi(e)) + [2\pi(e)(1 - \pi(e))\lambda^s \\
&\quad + \mu^s \cdot 2\pi'(e)(1 - 2\pi(e))]\nu'(t_1) = 0.
\end{align*}
These equations can be rewritten

\[ 2\psi'(u_0) = \lambda^w - \mu^w \frac{\pi'(e)}{1 - \pi(e)}, \]

\[ 2\psi'(u_{11}) = \lambda^w + \mu^w \frac{\pi'(e)}{\pi(e)}, \]

\[ 2\psi'(u_{10}) = \lambda^w - \mu^w \frac{\pi'(e)}{\pi(e)} - 2\mu^s \frac{\pi'(e)}{\pi(e)(1 - \pi(e))^3}, \]

\[ 2\psi'(u_2) = \lambda^w + \mu^w \frac{\pi'(e)}{\pi(e)} + 2\mu^s \frac{\pi'(e)}{\pi(e)(1 - \pi(e))^3}, \]

\[ \frac{1}{v'(t_0)} = \lambda^s - \mu^s \frac{2\pi'(e)}{1 - \pi(e)}, \]

\[ \frac{1}{v'(t_1)} = \lambda^s + \mu^s \frac{2\pi'(e)(1 - 2\pi(e))}{(1 - \pi(e))\pi(e)}, \]

\[ \frac{1}{v'(t_2)} = \lambda^s + \mu^s \frac{2\pi'(e)}{\pi(e)}. \]

Transfers follow the equations that are familiar from the principal-agent literature. Here, \( \mu^s > 0 \) because, if not, the supervisor's transfer would be decreasing in production and, in the normal case, we would be in the corner solution \( e^* = 0 \). Looking now at the equations defining the workers' trans-
fers, we see that \( u_{11} \neq u_0 \) and \( u_2 \neq u_{11} \). Therefore, transfers are not purely personalized. Moreover, we also see that \( u_{10} < u_0 \).

REFERENCES


