Monitoring and Collusion with "Soft" Information

Sandeep Baliga
Northwestern University

In the standard principal-supervisor-agent model with collusion, Tirole (1986) shows that employing a supervisor is profitable for the principal if the supervisor's signal of the agent's cost of production is "hard" (i.e., verifiable but hideable). Anecdotal evidence suggests that information is sometimes "soft" (i.e., unverifiable). We show that, in fact, it is profitable to employ a supervisor when information is "soft" even though the three parties can collude. Therefore, standard applications of the principal-supervisor-agent model to regulation and auditing have more scope than previously thought.

1. Introduction

We often observe an agent or supervisor who is employed by a principal to monitor other agents in situations where there is asymmetric information. A firm employs an auditor to monitor a manager's costs; a government employs a regulator to inspect a regulated firm; a firm employs a manager in part to monitor a worker. These situations can be thought of as vertical, three-tier principal-supervisor-agent hierarchies [see Antle (1982) for one of the first such models]. The supervisor is employed as he receives a signal that is helpful in designing an incentive scheme for the agent.

However, the supervisor's usefulness can be compromised if he and the agent collude to misrepresent their information. By collusion, we mean that the supervisor and agent can coordinate their messages to the principal and also make side payments. When regulators are captured, when auditors/directors work with management rather than for shareholders, is there indeed any use in employing a supervisor at all? Tirole (1986, 1992)[1] answers this question in a model where a principal uses a supervisor to screen two types of workers with different disutilities of effort. Despite collusion, Tirole (1986, 1992) shows

1. Tirole (1986) considers a model where both the supervisor and agent are risk-averse, there is unlimited liability and there are no transaction costs to side-payments. Tirole (1992) mainly uses a model where both the supervisor and agent are risk-neutral, there is limited liability and there can be transaction costs to side-payments. We will focus purely on the latter model.

that the principal is better off employing a supervisor when the supervisor's information about the worker is "hard" (roughly speaking this means verifiable but hideable) and there are transaction costs to side payments. The concept of hard information has also been used in the theoretical accounting literature (Antle, 1984; Ijiri, 1984; Kofman and Lawaree, 1993, 1996; Villadsen, 1995) and in the regulation literature (Laffont and Tirole, 1991a, b; 1992a, b).

But "soft" information is also an important environment to consider. Dalton (1959:31), in his study of the "Milo Fractionating Plant," reports that costs can be understated:

> Middle and lower officers learnt that an excellent cost record was directly related to their future. . . The more efficient heads were lauded, the others shamed. This scheme of reward and penalty led to ingenious distortions of records to more nearly approach ideal cost figures.

In Tirole's model, it is not possible for an auditor to understate a cost in the manner Dalton identifies. Dalton (1959:32–33) also reports instances where costs were exaggerated by creation of fictitious personnel or overstatement of costs of equipment. Even objective scientific data was manipulated.[2] This suggests that it is important to consider the case of soft information in the canonical models.

We show in the soft information counterpart of the Tirole (1992) model that *employing a supervisor is still useful for the principal even though information is soft and the three parties can collude*.

The main reason for employing a supervisor in the first place is that he receives a signal about the agent's cost of effort which the agent also knows. We show that the supervisor's report of his information can be cross-checked against the agent's. The potential difficulty is that collusion between the supervisor and agent might imply that they jointly misrepresent their information. This is prevented by sharing surplus in different ways for different cost levels: when they both report cost is low, the supervisor is given the surplus; when they both report costs are high, the agent is given any surplus; when their reports differ, neither is paid. Collusion is not a problem in such a mechanism.

Therefore, we show that standard results in the regulation and accounting literature are robust to the introduction of soft information even though collusion is possible. On a more speculative level, our analysis suggests that where there is a separation of duties within an organization but agents share soft information, it is possible for the principal to elicit their information if there are transaction costs to side-payments even if it is possible for these agents to collude. The

2. Dalton (1959:85) reports: "The solution in a series of vats was supposed to have a specific strength and temperature and a fixed rate of inflow and outflow in order to give desired qualities to the product. . . . [However], the solution was about triple the ideal strength, the temperature around ten degrees above normal, and the rate of flow double that expected. Yet the chemists sampled the solution at regular periods daily and by "graphite analysis" showed that all points met the laboratory ideal."

larger the set of agents that share information, the larger may be the transaction costs of collusion and the larger the gain to getting them to report on each other in the way we suggest.

In the next two sections, we present the model and the case of hard information. In the final section, we analyze a mechanism to elicit soft information.[3]

2. Principal-Supervisor-Agent Model

We utilize the canonical model reported in Tirole (1992). There are three parties: the principal, a supervisor, and an agent. The agent supplies 0 or 1 unit of a good to the principal. We assume that output is observable and verifiable. The agent has a privately known cost of supply β that takes on two values β^{H} and β_{L} with probability μ and $(1-\mu)$, respectively, where $\Delta\beta \equiv \beta^{\mathrm{H}} - \beta_{\mathrm{L}} > 0$. The surplus to the principal is S if the agent produces and 0 otherwise. We assume that

$$S - \beta^{\mathrm{H}} > 0. \tag{1}$$

That is, the principal wants even the high-cost agent to produce. Let $w \geq 0$ represent the expected wage of the agent and $x \in \{0, 1\}$ the quantity of the good produced.

The agent is risk neutral when utility is $U = w - \beta x$ and reservation utility zero. We assume

$$S - \beta^{\mathrm{H}} > (1 - \mu)(S - \beta_{\mathrm{L}}). \tag{2}$$

That is, the principal does not try to screen the worker's type without some extra information. This leads to the following result:

Claim 1 (Tirole, 1992). The principal's total cost of implementing his optimum is β^{H} when there is no supervisor.

The supervisor is also risk neutral and has income $s \geq 0$ and utility $V(s) = s$. His reservation utility is also zero. Assume that the agent can transfer income to the supervisor at a rate $k \in (0, 1]$ (i.e., if the agent transfers t pounds, the supervisor receives kt pounds). The supervisor learns a signal $\sigma \in \{\beta_{\mathrm{L}}, \varnothing\}$. If $\beta = \beta_{\mathrm{L}}$, then the supervisor learns it with probability α and learns nothing with probability $(1 - \alpha)$. If $\beta = \beta^{\mathrm{H}}$, then the supervisor learns nothing. While the information structure is unusual, it is actually the standard one utilized in the regulation and auditing literature and has proved very useful in that context.

The timing is as follows:

- Time 1: The agent learns his cost of effort β. The supervisor and the agent learn σ.
- Time 2: The principal offers the supervisor and agent a contract. This consists of messages to be sent to the principal and wage payments as a function of messages and output.

3. In recent work, Faure-Grimaud, Laffont, and Martimort (1998) analyze a model of a risk-averse auditor with soft information and endogenize the transactions costs of side payments.

Time 3: The supervisor offers the agent sign a side-contract consisting of messages, output, and a side payment.
Time 4: Contracts are implemented.

3. Hard Information

Tirole (1992) assumes the information the supervisor receives is hard: if the supervisor shows the signal $\sigma = \beta_{\text{L}}$ to the principal then the latter knows for certain that the agent's cost of effort is low. The supervisor's message space is $\{\sigma, \varnothing\}$. Therefore, if the supervisor receives no signal, he cannot claim that the agent's cost of effort is low.

The difficulty for the principal is that the supervisor can be bribed by the agent to hide a signal $\sigma = \beta_{\text{L}}$. As the agent is willing to pay $\Delta\beta$ to the supervisor to get him to hide his signal, the principal must pay $s = k\Delta\beta$ when he reveals it. This leads to the following result:

Claim 2 (Tirole, 1992). The principal's cost of implementing his optimum is $\mu\beta^{\text{H}} + (1-\mu)[\alpha(\beta_{\text{L}} + k\Delta\beta) + (1-\alpha)\beta^{\text{H}}]$ when information is hard.

A comparison of Claims 1 and 2 shows that when information is hard and there is some deadweight loss to transfers, it is useful to employ a supervisor. However, it may be the case that it is possible to falsify audit information so there is no such thing as convincing evidence about costs. Information is soft rather than hard. But then, in the mechanism above, even when the supervisor receives no signal he can claim that he receives a signal $\sigma = \beta_{\text{L}}$. With positive probability, $(1-\mu)(1-\alpha)$, the supervisor will get paid $k\Delta\beta$ when the agent produces output. As he gains from this deviation, the mechanism proposed above does not implement the principal's optimum when information is soft.

4. Soft Information

Suppose the principal designs the following mechanism. The supervisor and the agent simultaneously announce $m_M \in \{\beta_{\text{L}}, \varnothing\}$ and $m_A \in \{\beta_{\text{L}}, \varnothing\}$ at time 2. The agent decides whether to supply the good or not. The outcome function $g(m, x)$ is as follows: If $x = 0$ or if the supervisor and agent send different messages, the principal does not pay the supervisor or agent. If both announce $\varnothing$ and $x = 1$, the principal pays the agent β^{H} and the supervisor zero. If both announce β_{L} and $x = 1$, the principal pays the agent β_{L} and the supervisor $k\Delta\beta$. Let $w(m, x)$ and $s(m, x)$ be the agent's and the supervisor's expected gross payments for message profile m and output x.

In this model, there is an element of moral hazard as the agent can choose whether to supply the good or not. In state β given message profile m, the agent will prefer to produce an unit of output if and only if $w(m, 1) - \beta \geq w(m, 0)$. Let $x^*(m, \beta)$ be the optimal decision for the agent in state β for message profile m.

A strategy for the agent for state β and signal σ, $s_A(\beta, \sigma) = (m_A(\beta, \sigma), x(m, \beta))$, is a message $m_A(\beta, \sigma)$ and an output decision rule $x(m, \beta)$. A strategy for the supervisor for signal σ, $s_M(\sigma)$, is a message $m_M(\sigma)$. A strategy profile δ is a perfect Bayesian equilibrium if neither the supervisor nor the agent has an incentive to change his strategy at any information set given the strategy of the other and beliefs obey Bayes' rule whenever possible.[4] Notice that in any perfect Bayesian equilibrium the agent will produce output in state β and message profile m iff $x^*(m, \beta) = 1$.

The following strategy profile δ^* is perfect Bayesian equilibrium: Both parties announce the truth for all signals; the agent produces an output of 1 in state β after message profile m iff $x^*(m, \beta) = 1$. Notice that the equilibrium outcome for both signals is interim efficient (where interim efficiency accounts for the transaction cost of side payments). This suggests that the perfect Bayesian equilibrium is robust to collusion and we now in fact offer a definition of collusion-proofness.

First, we focus on supervisor and agent coalitions. Let $s(\delta \mid \sigma)$ be the supervisor's expected wage payment given the signal σ. A *collusive strategy* for the supervisor and the agent is a message profile m', an output x', and side-payment t' from the agent to the supervisor where $w(m', x') \geq t' \geq -s(m', x')$. Given a strategy profile δ, a collusive strategy $u = (m', x', t')$ for the supervisor and the agent is *acceptable for signal* σ if $s(m', x') + t' > s(\delta \mid \sigma)$ and $w(m', x') - x'\beta - t' > w([m_M(\sigma), m_A(\sigma, \beta)], x^*([m_M(\sigma), m_A(\sigma, \beta)], \beta)) - x^*([m_M(\sigma), m_A(\sigma, \beta)], \beta)\beta$ for some β which has positive probability given the signal σ.

A strategy profile δ is a *collusion-proof equilibrium* if it is a perfect Bayesian equilibrium and if there is no acceptable collusive strategy for the supervisor and the agent for any signal.

This is a strong notion of collusion-proofness. We could have demanded that a collusive strategy make *all* types that have positive probability better off rather than just some types. This makes collusion more difficult.[5] Therefore, if a perfect Bayesian equilibrium satisfies our notion of collusion-proofness, it certainly satisfies the weaker notion just suggested.

We claim the perfect Bayesian equilibrium δ^* is also a collusion-proof equilibrium. There is no incentive for any signal for the supervisor and agent to agree to a collusive strategy that involves announcing messages that do not match as then neither gets paid. The only potentially profitable collusive strategy for signal $\sigma = \beta_{\mathrm{L}}$ is to jointly announce β^{H}, for the agent to produce one unit of output and to make a monetary transfer at least $k\Delta\beta$ to the supervisor. But then, as the agent cannot be strictly better off, this is not acceptable for the supervisor and the agent. For the signal $\sigma = \varnothing$, if the state is β_{L}, the agent earns a rent of $\Delta\beta$ in equilibrium and cannot be better off in any collusive strat-

4. See Fudenberg and Tirole (1992) for a formal definition of a perfect Bayesian equilibrium.

5. Formally, given a strategy profile δ, a collusive strategy $u = (m', x', t')$ for the supervisor and the agent is *universally acceptable for signal* σ if $s(m', x') + t' > s(\delta \mid \sigma)$ and $w(m', x') - x; \beta - t' > w([m_M(\sigma), m_A(\sigma, \beta)], x^*([m_M(\sigma), m_A(\sigma, \beta)], \beta)) - x^*([m_M(\sigma), m_A(\sigma, \beta)], \beta)$ for all β which have positive probability given the signal σ. Notice that, given some strategy profile, a universally acceptable collusive strategy for some signal is also acceptable for that signal. A strategy profile δ is a *weakly collusion-proof equilibrium* if it is a perfect Bayesian equilibrium and if there is no universally acceptable collusive strategy for the supervisor and the agent for any signal. Notice that a collusion-proof equilibrium is also weakly collusion-proof.

egy where both he and the supervisor announce β_L. Therefore there is clearly no acceptable collusive strategy for the signal $\sigma = \varnothing$ either. Hence δ^* is a collusion-proof equilibrium.

However, we have not yet allowed for collusion between the principal and his employees. The mechanism above can prevent such collusion with the following change: if the agent and supervisor send different messages, then confiscate a large amount of money from the principal.[6] Therefore, the principal never has an incentive to bribe the supervisor or the agent for any signal in such a way that their messages do not match. We need only consider if he has any incentive to make both the agent and supervisor change their messages. When the signal is $\varnothing$, the principal cannot possibly reduce the cost of implementing his optimum further. When the signal is β_L, bribing both the agent and the supervisor to change their messages to $\varnothing$ cannot make the principal better off. This discussion is summarized in the following result:

Proposition. Suppose information is soft. The principal's cost of implementing his optimum is $\mu\beta^H + (1-\mu)[\alpha(\beta_L + k\Delta\beta) + (1-\alpha)\beta^H]$, the same as the cost when information is hard.

Therefore the principal-supervisor-agent model is robust to the introduction of soft information. The standard models in the theoretical accounting and regulation literatures (which, to be sure, are special) are not compromised by the soft information to which Dalton (1959) refers.

Acknowledgments

Many thanks to Patrick Bolton, Oliver Williamson, and two referees for comments. Any errors are mine.

References

Antle, R. 1982. "The Auditor as an Economic Agent," 20 *Journal of Accounting Research* 503–27.

Antle, R. 1984. "Auditor Independence," 22 *Journal of Accounting Research* 1–20.

Dalton, M. 1959. *Men who Manage*. New York: John Wiley & Sons.

Faure-Grimaud, A., J.-J. Laffont, and D. Martimort. 1998. "A Theory of Supervision with Endogenous Transaction Costs," mimeo, Toulouse.

Ijiri, Y. 1971. "A Defence of Historical Accounting," in R. B. Sterling, ed., *Asset Valuation and Income Determination*. Kansas City: University of Kansas Press.

Kofman, F., and J. Lawaree. 1993. "Collusion in Hierarchical Agency," 61 *Econometrica* 629–56.

Kofman, F., and J. Lawaree. 1996. "On the Optimality of Allowing Collusion," 61 *Journal of Public Economics* 383–407.

Laffont, J.-J., and J. Tirole. 1991a. "Auctions and Favoritism," 9 *International Journal of Industrial Organization* 9–42.

Laffont, J.-J., and J. Tirole. 1991b. "The Politics of Government Decision-Making: A Theory of Regulatory Capture," 106 *Quarterly Journal of Economics* 1089–127.

Laffont, J.-J., and J. Tirole. 1992a. "Cost-padding, Auditing and Collusion," 25–26 *Annales d'Economie et Statistique* 205–26.

6. Tirole (1986) also utilizes such a construction in the proof of his Proposition 6 to prevent collusion between the principal and his employees.

Laffont, J.-J., and J. Tirole. 1992b. “Should Governments Commit?,” 36 *European Economic Review* 345–53.

Tirole, J. 1986. “Hierarchies and Bureaucracies: On the Role of Collusion in Organizations,” 2 *Journal of Law, Economics, & Organization* 181–214.

Tirole, J. 1992. “Collusion and the Theory of Organizations,” in J.-J. Laffont, ed., *Advances in Economic Theory*, vol. I. Cambridge: Cambridge University Press.

Villadsen, B. 1995. “Communication and Delegation in Collusive Agencies,” 19 *Journal of Accounting and Economics* 315–44.