

Explicit and Implicit Incentives for Multiple Agents

By Jonathan Glover

Contents

1	Introduction	2
2	Adverse Selection	14
2.1	Second-Best Contracts	14
2.2	The Multiple Equilibrium Problem and “Simpler” Mechanisms	20
2.3	Robust Mechanisms	25
3	Moral Hazard	31
3.1	Second-Best Contracts	31
3.2	The Multiple Equilibrium Problem and “Simpler” Mechanisms	34
3.3	Optimal Use of Additional Performance Measures	36
3.4	Bailouts and Coordinated Moral Hazard	37
3.5	Robust Mechanisms	40
4	Implicit Contracts	42
4.1	Two-Period Relational Contracts Between Agents	42
4.2	Infinite Horizon Relational Contracts Between Agents	47
4.3	Relational Contracts with and Between Agents	50

5 Conclusion	53
5.1 “Simpler” Mechanisms	54
5.2 Robust Mechanisms	55
5.3 Implicit Contracts	57
A Appendix	60
Acknowledgments	64
References	65

Foundations and Trends® in
Accounting
Vol. 7, No. 1 (2012) 1–71
© 2012 J. Glover
DOI: 10.1561/1400000020



Explicit and Implicit Incentives for Multiple Agents

Jonathan Glover

*Tepper School of Business, Carnegie Mellon University, Pittsburgh,
PA 15213, USA, jglover@andrew.cmu.edu*

Abstract

This monograph presents existing and new research on three approaches to multiagent incentives. The goal of all three approaches is to find theories that better explain observed institutions than the standard approach has.

1

Introduction

Mechanism design theory treats institutions as endogenous games. In accounting, research on mechanism design has focused largely on principal–agent models. In the standard principal–single agent model, the principal offers the agent an incentive contract and then leaves the agent with what is essentially a decision problem. Yet, even simple principal–agent models can produce complicated optimal contracts. For example, an important result from moral hazard models with a risk-averse agent is that all informative variables will be incorporated into the optimal contract (Holmstrom, 1979).¹ A performance measure is informative if its conditional probability distribution depends on the agent’s action, where the conditioning is on all performance measures already incorporated into the contract. Holmstrom’s result provided a theory of relative performance evaluation (Antle and Smith, 1986) and refined the accountant’s traditional notion of controllability to “conditional controllability” (Antle and Demski, 1988). The broader information content school of accounting theory has developed a better

¹Holmstrom’s (1979) informativeness condition is both necessary and sufficient for it to be optimal to incorporate an additional performance measure into a contract when the agent is risk averse and the incentive compatibility constraint binds.

(more nuanced) understanding of a wide variety of managerial and financial accounting practices (e.g., Demski, 2010; Christensen and Demski, 2003).

With a large number of informative variables, which seem inevitable in practice, the optimal contract Holmstrom predicts would be overwhelmingly complex. Even when there is only a single variable to contract on, the optimal contract can be extremely sensitive to the underlying details of the environment. For example, with a risk-neutral agent, the optimal contract can have the principal making an extremely large payment to the agent with an extremely small probability. If the probabilities are different than assumed, the principal may end up paying the agent much more than is required or fail to motivate the agent to take the action she intends. Real-world incentive contracts seem less fine-tuned to the environment and more robust.

When a principal contracts with multiple agents, even more extreme results emerge. For example, in models of capital budgeting with multiple risk-neutral agents who operate in correlated cost environments, the optimal contract prescribes some payments that are arbitrarily large and others that are arbitrarily small (negative) as the correlation becomes small. These arbitrarily large and small payments allow the principal to extract all of the agents' information rents and obtain the first-best solution as long as there is any correlation in project returns. The ease of achieving first-best payoffs and the knife-edged nature of the optimal contract make it suspect as an explanation of anything we see in practice (Cremer and McLean, 1988).

With risk-averse agents, the contract is less knife-edged in response to the risk premium associated with imposing risk on the agents but presents another problem. The optimal Bayes-Nash incentive compatible contract (the second-best solution) typically creates multiple equilibria in the agents' subgame, and the agents may find tacitly colluding on an equilibrium other than the one the principal intends them to play appealing (Demski and Sappington, 1984; Mookherjee, 1984). That is, the second-best solution may induce excessive (from the principal's perspective) coordination.

This monograph presents research on three themes related to multiagent incentives, taking the view that developing a better

4 *Introduction*

understanding of multiagent incentives is central to developing a better understanding of observed institutions. The organizing theme is multiple equilibria created by the use of the Bayes–Nash solution to the multiagent contracting problem. First, in preventing tacit collusion, confession is an alternative to ratting that allows for less demanding behavioral assumptions than Bayes–Nash, while approximately implementing the second-best solution (Glover, 1994). Second, optimal robust contracts designed to deal with a variety of settings are qualitatively similar to the standard optimal contracts when the variety is small and qualitatively different than the standard ones when the variety is large. When the variety is large, individual rather than relative performance evaluation is optimal in moral hazard settings, and procurement contracts similar to observed second-price procurement auctions emerge as optimal in adverse selection (Arya et al., 2009). Such contracts are not subject to the tacit collusion problem by virtue of providing dominant strategy incentives. Third, in repeated settings, collusion can be turned into cooperation (implicit contracting between the agents that benefits the principal) by using aggregate performance measures to motivate mutual monitoring by the agents (Arya et al., 1997a). The monograph surveys existing research on these three themes (with a fairly narrow focus on my own research) and presents a few new results. Rather than presenting each of the models employed in the existing papers, two basic models are used — one for adverse selection and one for moral hazard. The goal is to present the results as simply as possible.

One approach to dealing with unwanted coordination is to make obedient behavior a dominant strategy (e.g., Demski and Sappington, 1984). Using a two-agent model of moral hazard, I expand on this approach, allowing for the production of additional (monitoring) information that provides information about individual efforts. The additional information can be thought of as produced by an audit of each agent’s action. The second-best solution, which motivates each agent to “work” rather than “shirk” given the other agent is playing “work,” itself incorporates this additional information. Under the revised second-best solution, each agent’s pay now also depends on the audit of his own effort but not on the audit of the other agent’s

effort, since the audit of the other agent's effort is not informative in the sense of Holmstrom (1979).

If instead the optimal contract has to eliminate an equilibrium that has both agents playing “shirk” instead of both playing “work” by making “work” a dominant strategy, the optimal contract incorporates all information. In particular, each agent's pay is highest when the audit of his own effort indicates he has played “work” and the audit of the other agent's effort indicates he has played “shirk.” This is a way of turning up the power of incentives when the audit of the other agent's effort indicates the agents may be playing the bad equilibrium. It is optimal to use an uninformative signal, because doing so is the optimal way to prevent collusion (without adding self-reporting). As far as I know, this is a new result. Put differently, although the individual signals are uncorrelated, the possibility of tacit collusion creates an endogenous correlation that optimal dominant strategy contracts incorporate but the standard optimal second-best Nash contracts ignore.

The mechanism design literature points us in a different direction: we augment the second-best solution by adding (costless) self-reporting. This new self-reporting is used to eliminate unwanted equilibria without creating new equilibria or changing the equilibrium payoffs (e.g., Ma, 1988; Ma et al., 1988; Mookherjee and Reichelstein, 1990). These augmented mechanisms are typically complex, for example, employing infinite message spaces when the underlying type space is binary. The typical “tail-chasing” mechanisms also exploit a weakness of the Nash equilibrium concept — that best responses are not always well defined. Arguably, these mechanisms without well-defined best responses are of limited use in understanding actual institutions (Jackson, 1992). The focus of the implementation literature has been on what can and cannot be implemented, not the form of the implementing mechanisms.

The monograph presents simpler mechanisms than those usually used to eliminate unwanted equilibria (e.g., Glover, 1994). These simpler mechanisms assume less demanding behavioral assumptions — two rounds of iteratively removing strictly dominated strategies — and employ smaller message spaces than the standard ones but achieve only approximate implementation of the second-best solution. Hence, the approach represents an arbitrarily small deviation from the standard

approach of searching for institutions in optimality. As examples of practices that resemble the mechanisms, budgeting (and budget padding in particular) in adverse selection (Arya and Glover, 1996) and management forecasts in moral hazard (Arya and Glover, 1995) can be viewed as providing opportunities for off-equilibrium confessions.² If the agents have complete information, even the first-best solution can be exactly implemented via two rounds of iteratively removing strictly dominated strategies in a general principal–multiagent model of adverse selection (Arya et al., 2000b). The approach relies on the reports of other agents in determining any one agent’s equilibrium allocation, while providing each agent with the opportunity to challenge what others say about him. A challenge is appealing to an agent if and only if other agents are lying about him. Put in terms familiar to accountants, when two managers or a manager and an auditor can both verify something (e.g., the historical cost of an asset), the mechanism design literature suggests that information should be relatively easy to elicit in principal–multiagent models.³

The principal may also be a player in the game, beyond her role in designing incentive contracts. For example, the principal may be able to bail the agents out when the outcome would otherwise be disastrous for her. Arya and Glover (2006) study a potential bailout by a principal that is more likely to occur when early signals indicate both agents’ projects are likely to fail. A familiar idea in banking regulation is that bailouts create moral hazard.

Repeated intervention subverts the incentives that are the moving force of market behavior. Bailouts obviate the hard choices — default or reform for troubled borrowers; sound lending judgments or failure for investors — and substitute a free ride on taxpayers in the Group of Seven leading industrial nations. Capital markets have learned that there is an implicit

²The principal commits not to punish an agent who confesses. To quote Bassanio from Shakespeare’s *Merchant of Venice*, “Promise me life, and I’ll confess the truth.”

³An important caveat is that most of the mechanism design literature (including Arya et al., 2000b) confines attention to single-period settings.

International Monetary Fund guarantee for large emerging market borrowers and that risk premium can be collected while avoiding the risk (Lerrick and Meltzer, 2001).

Similar arguments against bailouts were made in the wake of the more recent subprime mortgage crisis. The point here is related but different: a bailout that is more likely when multiple institutions are likely to fail leads to *coordinated* moral hazard. The potential of a bailout creates an endogenous correlation in the agents' environments. Even if taking desirable actions (working, diversifying, due diligence in credit evaluations, etc.) is a Nash equilibrium, the possibility of a bailout may lead the managers to take on coordinated undesirable actions in order to make it more likely they will be bailed out.

Relative performance evaluation is a natural solution. In the case of the banking crisis, one version of relative performance evaluation has the surviving banks receiving the assets of the failed banks at a bargain price (Acharya and Yorulmazer, 2007). In an earlier paper, Arya and Glover (2006) study a principal–two-agent model of moral hazard subject to a bailout by the principal and explore the role of information system design as a commitment device designed to limit bailouts and, hence, unwanted coordination.⁴

Returning to the critique that the traditional optimal mechanisms seem overly fine-tuned to the environment:

This brings me to a point I wish to emphasize: The optimal trading rule for a direct revelation game is specialized to a particular environment. For example, the rule generally depends on the agents' probability assessments about each other's private information. If left in this form, therefore, the theory is mute on one of the most basic problems challenging theory. I refer to the problem of explaining the prevalence of a few simple trading rules in most of the commerce

⁴The working paper version from 2001 provides a more general, although still highly stylized, analysis and is available online.

conducted... The rules of these markets are not changed daily as the environment changes; rather they persist as stable, viable institutions. As a believer that practice advances before theory, and that the task of theory is to explain how it is that practitioners are (usually) right, I see a plausible conjecture: These institutions survive because they employ trading rules that are efficient for a wide class of environments (Wilson, 1987, pp. 36–37).

The second approach the monograph takes is to focus on the simplicity of the second-best solution itself by requiring the mechanism be designed early, before a particular application (environment) arises. The optimal robust mechanism is found taking expectations over the potential environments. The early design assumption is equivalent to blocked communication — the mechanism is not allowed to be a menu of mechanisms that are later fine-tuned to the actual environment based on the agents' communication about the environment once the particular environment arises. Optimal robust mechanisms help rationalize, for example, second-price auctions (Arya et al., 2009). The auction model can be converted into a capital budgeting setting with mutually exclusive projects by changing signs. In the capital budgeting version of the model, the second-price auction can be reinterpreted as relative project ranking, which is a common means of rationing resources in organizations. Second-price auctions provide dominant strategy incentives. Hence, the result provides a Bayes–Nash foundation for dominant strategy mechanisms. Other approaches to the robustness problem (e.g., Bergemann and Morris, 2005, 2008) take larger departures from the Bayesian framework, focusing on, for example, robustness to higher order beliefs.⁵

⁵ Many of the implementing mechanisms used to prove sufficiency results in the recent literature on robust mechanisms are subject to the Jackson critique, which calls into question the notions of robustness being employed. Studying specific settings and robust mechanisms for those settings seems more likely to yield fruitful positive results and insights into observed institutions. So far, the robustness literature's most interesting results are negative ones.

The important point in Arya et al. (2009) is that the optimal mechanisms are qualitatively similar to the standard ones when robustness is a small concern but are qualitatively different from the standard ones when robustness is a large concern, although even a small robustness concern can convert a nonunique solution into a unique one.

I present a new result on robust contracts for moral hazard. Robust contracts designed by two principals of two different firms do not use relative performance evaluation when the standard optimal contracts would. The principals each understand the production technology their own agent operates but have less information about the technology operated in the competing firm.

Casting the firm as a principal–multiagent model in which the principal provides all incentives via explicit contracts is, at best, an abstraction of a broader relationship. As Sunder (1997) writes, the firm is “an arena in which self-motivated economic agents play by mutually agreed upon or implied rules to achieve their respective objectives.” The comparative advantage of a firm over other institutional arrangements is in enforcing implicit contracts, since the courts can enforce explicit contracts. Repeated relationships and multiple equilibria (in the continuation game) are an essential part of implicit contracts, since self-enforcing punishments agents can impose on each other are needed to make their implicit promises credible. That is, instead of viewing multiple equilibria as undesirable, the principal can use multiple equilibria to her advantage.

The third approach this monograph studies is to incorporate repeated play and implicit (relational) contracting among the agents and between the agents and principal. Relational contracting and mutual monitoring among the agents create a role for joint performance evaluation, even when the joint performance measure is an aggregation of individual performances measures that could be contracted on individually (Arya et al., 1997a; Che and Yoo, 2001).

When the relationship is repeated indefinitely, whether the agents’ actions are strategic complements or strategic substitutes is important. The agents’ actions are strategic complements if each agent’s marginal benefit of increasing his own action is increasing in each other agent’s action. An example of a setting in which individual efforts

are strategic complements is an interdisciplinary project — individual effort is most productive when the other team members are also working hard. An example of a setting in which actions are strategic substitutes is a project for which the agents' actions are interchangeable and there are decreasing returns to total effort. A strategic substitutability limits the gains to mutual monitoring and cooperation, because the agents are tempted to collude on taking turns doing the work (think of group projects in a classroom setting) unless the incentives are high powered.

Under a twofold repetition of the relationship, mutual monitoring is always optimal in the first period. Multiple equilibria are created in the second period that the agents use to enforce cooperation in the first period. If individual performance measures are available, only the sum is used in determining compensation in the first period, since this is the most efficient way to provide group incentives for joint working over joint shirking. In the second period, a mix of aggregate and individual performance evaluation is used. Aggregation is inefficient in providing second-period incentives, so just enough aggregation is used to provide the punishment needed to enforce first-period cooperation. In the two-period setting, the turn-taking collusion problem does not arise. The second-period incentives must be high powered enough to provide individual incentives for working, since the second period is the final period.

Mutual monitoring can be viewed as an alternative to the confession (and other self-reporting) mechanisms discussed earlier. The implicit contracting approach casts accounting and explicit contracting as means of setting the stage for (decentralized) implicit contracting rather than an all-encompassing (centralized) source of information and contracting. Earlier approaches to mutual monitoring (e.g., Itoh, 1993; Tirole, 1986) assumed agents could write explicit side contracts with each other, often describing explicit side contracting as an abstraction of a repeated relationship with implicit side contracting. The implicit side contracting approach is relatively underexplored (particularly finitely repeated implicit contracting) and has the potential to yield new insights into observed practices (e.g., the evolution of incentives over a manager's tenure).

Enforcement is a more serious issue for side contracts than for ordinary contracts. If collusion poses a threat to an organization, the latter may stipulate in its grand contract among members that side contracts or some forms of potentially verifiable side transfers are prohibited. Indeed, organizations routinely do so... If, as is often the case, repeated interaction is indeed what enforces side contracts, the second approach [of modeling repeated interactions] is clearly preferable because it is more fundamentalist; it takes a more agnostic view of whether gains from trade are realized within groups, and in doing so, it unveils an important control variable affecting the realization of collusion (Tirole, 1992, pp. 155–156).

It is an empirical question whether various groups of managers are best modeled using models of individual incentives (e.g., Holmstrom, 1979) or models of group incentives (e.g., Itoh, 1993; Arya et al., 1997a; Che and Yoo, 2001). In studying the impact of pay-for-performance dispersion on top executives' incentives, Bushman et al. (2012) use the Arya et al. (1997a) and Che and Yoo (2001) models to motivate studying the role team tenure has in mitigating poor performance. They speculate that the poor performance they observe is caused by individual free-riding. Their results are consistent with team tenure reducing free-riding.

In another recent paper, Li (2012) studies structural models of the incentives provided to top executives, using the consistency of the risk-aversion estimates to evaluate the ability of the models to explain observed incentives. She finds the most consistent estimates for a group incentive model similar to Arya et al. (1997a) with less consistent estimates from an individual incentive model similar to the Holmstrom (1979) model. Both of these models perform better than another team-based model that allows for explicit side contracts and transferable utility (similar to one of the models in Itoh, 1993).

Li (2012) raises intriguing questions about the existing literature on executive compensation, which criticizes executive compensation

practice for not being more consistent with models of optimal individual incentives (e.g., Jensen and Murphy, 1990; Bebchuk and Fried, 2006). Perhaps, it is time for the executive compensation literature to devote more attention to models of group incentives.

If contracting is limited to non-verifiable variables (e.g., subjective assessments of agents' actions), then the principal's contract with the agents is also an implicit one. If the agents observe each other's actions, then there is room for implicit contracting with and between agents. With only non-verifiable performance measures, bonus pools emerge as an optimal response to the principal's limited ability to make commitments but only as an extreme form of the optimal relational contract when the discount rate becomes extremely large and the model is essentially of a single-period setting (Glover and Xue, 2012). When the discount rate is small, the optimal contract is a group-based one that fosters implicit contracting and mutual monitoring among the agents as in Arya et al. (1997a). As the discount rate is increased, group-based pay becomes infeasible, and the principal uses a mix of individual and relative performance evaluation. The feature of bonus pools that has the principal rewarding the agents for poor performance emerges sooner than one might expect. The reason is that relative performance evaluation encourages the agents to collude on taking turns working. Pay for bad performance is used to mitigate this effect of relative performance evaluation — to make the agents' payoffs strategically independent instead of creating a strategic substitutability. (In one-shot games, strategic independence is a particular form of dominant strategy incentives.)

Baldenius and Glover (2012) study exogenous bonus pools in an indefinitely repeated setting in which there are non-verifiable individual measures that cannot be explicitly contracted on and a verifiable (objective) measure of team performance (e.g., firm-wide earnings) that can be contracted on. The strategic complementarity or substitutability of the agents' actions is again key, this time in the team performance. Bonus pools perform at their best when the agents' actions are closest to strategically independent. With a large strategic complementarity, it can be better to make the bonus pool independent of the team

performance measure instead of having the bonus pool be increasing in the team performance measure.⁶

The remainder of the monograph is organized into four sections. Sections 2 and 3 study explicit contracting. I begin with adverse selection in Section 2 and turn to moral hazard in Section 3. Section 4 studies implicit contracting. Section 5 concludes by discussing additional managerial and financial reporting (regulation) applications and potential extensions.

⁶In fact, in the case of a large strategic complementarity, it is optimal to make the bonus pool a decreasing function of the team performance measure, but this may not be a practical solution, since understating performance (or destroying output) is often possible.

2

Adverse Selection

2.1 Second-Best Contracts

A classic model of adverse selection is Antle and Eppen (1985). Their analysis develops a rationale for observed features of capital budgeting within firms — in particular, for organizational slack and capital rationing (imposing a hurdle rate for the required return on projects that exceeds the firm’s cost of capital). Antle and Eppen show that these practices can be optimal because of incentive considerations. In particular, capital rationing emerges as an optimal means of trading off efficient production and information rents the agent receives because of his private information.¹

Arya et al. (1996) extend the principal–single agent model of Antle and Eppen (1985) to include a second agent and show that relative project ranking, a common means of capital rationing, can be optimal

¹Antle and Fellingham (1990) studies a two-period extension of the Antle and Eppen (1985) model and finds that history-dependent contracts (contracts with memory) are optimal, despite their assumption of serially uncorrelated project returns. In Arya and Glover (2001), it is optimal to delay and bundle project decisions when the projects are iid, because bundling reduces the information asymmetry between budget headquarters and the manager. Antle and Fellingham (1997) surveys extensions of the Antle and Eppen (1985) model.

for informational reasons. Relative project ranking is used because of what one proposal indicates about the probability distribution of possible net present values (NPVs) of other projects, not because of a shortage of capital.

Consider a firm consisting of a profit-maximizing headquarters (principal) and two division managers (agents). The division managers, A and B , are risk neutral. Each manager is responsible for providing information to headquarters about potential projects.

Two features create contracting frictions: (i) each division manager has private information and (ii) each manager must have a non-negative wealth level at the end of the period (as long as he is truthful). Division managers do not have the ability to (directly) borrow funds from the capital markets.

The sequence of events is as follows: first, each division manager privately learns the cost of production for his project (which is equivalent to the rate of return); second, headquarters offers each manager a take-it-or-leave-it contract governing the funding of projects; and third, the managers submit budgets to headquarters.

Each investment opportunity earns a constant rate of return. In order to produce a cash inflow of x^i in one year, the cost is $c^i x^i$ now. c^i is the required investment cost. This linear structure sets up a linear programming problem. In order to ensure the existence of a solution, capacity is constrained: x^i can be at most X^i . Assume $c^i \in \{c_L, c_H\}$, where $c_L < c_H$. For simplicity, assume the discount rate used to present value cash flows is 0, i.e., the discount factor is 1. Both low- and high-cost projects have a positive NPV: $c_L < c_H < 1$.

Manager i privately observes c^i prior to contracting with headquarters and submits a budget \tilde{c}^i after he contracts with headquarters. Denote by $p_{km}, k, m = L, H$, the common knowledge prior probability $c^A = c_k$ and $c^B = c_m$.² To keep things simple, assume the managers are ex ante identical so that $p_{km} = p_{mk}$. Whenever it does not

²Arya et al. (2000a) study a model in which the probabilities depend on the manager's project search. Because contracting is delayed until the capital budgeting stage, coarse monitoring, coarse self-reporting, late monitoring, and/or late self-reporting can be optimal as a way for headquarters to commit to let the manager earn slack, which motivates the manager's project search. See also Dutta and Fan (2012).

cause confusion, superscripts will be suppressed. Denote by p_k the marginal probability $c^i = c_k$. Assume the managers' environments are either uncorrelated or positively correlated: $p_{LL}p_{HH} - p_{LH}p_{HL} \geq 0$.

The contracts offered by headquarters consist of budget-contingent transfers. Manager i 's contract specifies: (1) the transfer, denoted t_{km} , headquarters makes to manager i at the beginning of the period and (2) the cash inflow, denoted x_{km} , manager i must return to headquarters at the end of the period if budgets of $\hat{c}^A = c_k$ and $\hat{c}^B = c_m$ are submitted. In order to produce x_{km} , manager i must invest $c^i x_{km}$. Since c^i is privately observed by manager i , only he knows the actual investment needed to produce x_{km} . Any difference between headquarters' transfer to manager i and the actual investment needed to produce the cash inflow amount required by the contract, the project's slack of $t_{km} - c^i x_{km}$, is consumed by manager i . Headquarters retains the residual $x_{km} + x_{mk} - t_{km} - t_{mk}$. The manager maximizes expected slack, while headquarters maximizes her expected residual.

Headquarters' contracting problem is to maximize the expected residual subject to the following constraints. First, the individual rationality (IR) constraints require the contract be sufficiently attractive to each manager. Whether manager i 's cost is low or high, the contract must provide manager i with at least his reservation utility, denoted \bar{U} . Second, the Bayes–Nash incentive compatibility (IC) (truth-telling) constraints ensure each manager has incentives to truthfully budget (report) his project's cost, given the other manager's budget is truthful. Imposing these constraints simplifies the search for the optimal contract and, by the Revelation Principle (Myerson, 1979), is without loss of generality. That is, any equilibrium outcome of any mechanism can also be achieved under truth-telling by a direct mechanism for which truth-telling is an equilibrium. Third, the bankruptcy (B) constraints require the contract provide each manager with nonnegative slack if he is being truthful. That is, all funds for investment must be provided by headquarters and, as long as a manager is being truthful, headquarters cannot require that he produces a return on these investment funds that is greater than the rate of return (implied by the cost) he reported. Fourth, the output feasibility (OF) constraints require that the scale of production (the cash inflow from the

project) be less than or equal to X . Also, all choice variables must be non-negative (NN).

Headquarters' contracting problem is formalized in the Linear-Program(AS-RN).

$$\underset{x,t}{Max} E[x_{km} + x_{mk} - t_{km} - t_{mk}] \quad (\text{AS-RN})$$

s.t.

$$E_{m=L,H}[t_{km} - c_k x_{km} | c_k] \geq \bar{U}, \quad \forall k \quad (\text{IR})$$

$$E_{m=L,H}[t_{km} - c_k x_{km} | c_k] \geq E_{m=L,H}[t_{nm} - c_k x_{nm} | c_k], \quad \forall k, n \quad (\text{IC})$$

$$t_{km} - c_k x_{km} \geq 0, \quad \forall k, m \quad (\text{B})$$

$$x_{km} \leq X, \quad \forall k, m \quad (\text{OF})$$

$$x_{km}, t_{km} \geq 0, \quad \forall k, m \quad (\text{NN})$$

The Revelation Principle is used in formulating the program: communication is confined to a cost report of c_L or c_H , and truthful reporting is motivated. So, the subscripts in the objective function refer to the actual costs. For simplicity, assume $\bar{U} = 0$, which ensures that the (IR) constraints do not bind.

The model is easily adapted to a variety of other settings. Team production can be incorporated by adding the constraint: $x_{km} = x_{mk}$. Procurement of a single indivisible item can be incorporated by requiring $x_{km} + x_{mk} = 1$, where x_{km} denotes the probability of agent i being awarded the contract. Getting from procurement to an auction is a matter of changing signs.

An important benchmark is obtained by removing the bankruptcy constraint from the program. *s denote optimal contracts.

Proposition 2.1. In Program (AS-RN) without the bankruptcy constraints (B), the first-best solution can be obtained in dominant strategies by setting: $x_{km}^* = X$ for all km ,

$$t_{LL}^* = t_{HL}^* = \left(\frac{\Pr(c_H|c_H)c_L - [1 - \Pr(c_L|c_L)]c_H}{\Pr(c_H|c_H) + \Pr(c_L|c_L) - 1} \right) X,$$

and

$$t_{LH}^* = t_{HH}^* = \left(\frac{\Pr(c_L|c_L)c_H - [1 - \Pr(c_H|c_H)]c_L}{\Pr(c_H|c_H) + \Pr(c_L|c_L) - 1} \right) X.$$

Proof. See the appendix. □

While the Program (AS-RN) imposed only Bayes–Nash incentive constraints, the solution given in Proposition 2.1 (which is not the only solution) also provides dominant strategy incentives. Truth-telling is a dominant strategy for each manager but in a trivial way, since his report is used only in determining the other manager’s allocation. Nevertheless, a small tweak in the optimal contract can produce truth-telling as a unique strict dominant strategy. When a manager reports his cost is high, reduce production by an arbitrarily small amount, say ε , and reduce the payment by an amount, say $\varepsilon(c_L + c_H)/2$, that makes the reduction in production and payment desirable if and only if the manager’s cost is actually c_H .

Proposition 2.1 is essentially a special case of the main results of Cremer and McLean (1988) and Riordan and Sappington (1988). With more than binary costs, a spanning condition has to be satisfied in order to ensure the managers’ information rents can be extracted (the first-best can be obtained). Relative to the early, mostly negative, results on dominant strategy implementation (e.g., Gibbard, 1973; Satterthwaite, 1975) Proposition 2.1 is remarkable. The permissiveness of the result raises suspicion of the theory as a source of insights into observed institutions.

In “nearly all” auctions, the seller should be able to extract the full surplus, which implies that asymmetry of information between buyers and sellers should be of no practical importance. Economic intuition and informal evidence (we know of no way to test such a proposition) suggest this result is counterfactual... Costly information gathering, not explicitly modeled in auction problems, may result in less profitable but vastly simpler auctions being used in practice.

(Cremer and McLean, 1988, p. 1254)

Proposition 2.1’s mechanism is extremely fine-tuned to the environment. As the correlation goes to 0, t_{LL} and t_{HL} go to negative infinity and t_{LH} and t_{HH} go to positive infinity. One way of assuming “costly

information gathering” is to assume the principal cannot obtain all of the details (e.g., the exact correlation) before designing the incentive contract/mechanism. We will come back to this point later in the monograph when robustness is studied.

Once the bankruptcy (B) constraints are included, the optimal mechanism is simpler and, arguably, less suspicious as a description of observed institutions. The optimal contract turns out to be an intuitive variation of the single-agent solution.

With a single manager, the solution is one of two contracts: Rationing or Slack. Under Slack, all projects are funded as if they are high-cost projects. Under Rationing, only low-cost projects are funded, and the manager earns no information rents. The tradeoff is the same as in other models of adverse selection: the principal curtails production for all but the highest type (lowest cost) in order to reduce the information rents that need to be paid to the agent because of his private information. The linear model makes the tradeoff stark.³

With two managers, the only modification to the single-manager solution is that whether Slack or Rationing is used for one manager may depend on what the other manager reports (Arya et al., 1996). The third optimal contract (which combines the others) can be interpreted as relative project ranking, since each division’s project is turned down only when he reports his own cost is high and the other manager reports his cost is low.

Proposition 2.2. The solution to Program (AS-RN) can be characterized so that it provides the division managers with dominant strategy incentives to report truthfully. This solution is Rationing, Slack, or Relative Project Ranking, where Relative Project Ranking is:

$$\begin{aligned} x_{LL}^* &= X, & x_{HL}^* &= 0, & t_{LL}^* &= c_L X, & t_{HL}^* &= 0 & \text{and} \\ x_{LH}^* &= x_{HH}^* = X & \text{and} & & t_{LH}^* &= t_{HH}^* = c_H X. \end{aligned}$$

Proof. See Arya et al. (1996). □

³With more than binary cost realizations, the optimal contract is a hurdle rate (cost) contract. If the reported cost is at or below the cutoff c^* , the project is funded as if it were a c^* cost project. If the reported cost is higher than c^* , the project is rejected.

Again, the solution is not unique. The project acceptance decisions are unique, but the payments are not. However, any of the other solutions are equivalent to Proposition 2.2's characterization in the sense that the payoffs to all parties are the same.

What features of the (AS-RN) model make it possible to equivalently characterize any Bayes–Nash optimal incentive compatible allocation as one that also satisfied more demanding dominant strategy incentive constraints? The linear cost structure is unimportant. The payments rather than the production levels are adjusted to convert the Bayes–Nash incentives into dominant strategy incentives. The additively separable utility function with risk neutrality in wealth is important. Also, correlated types get in the way unless there are only two possible types.

With uncorrelated types, agent risk neutrality, and other fairly standard assumptions, any allocation rule that satisfies Bayes–Nash incentive constraints can be equivalently implemented (keeping the payoffs the same) in dominant strategies (Mookherjee and Reichelstein, 1992). In another analysis of capital budgeting, Baldenius et al. (2007) study a setting that satisfies the Mookherjee and Reichelstein (1992) conditions. They show that the nature of the investment — shared versus competing — determines the nature of the hurdle rate's deviation from the firm's cost of capital.

2.2 The Multiple Equilibrium Problem and “Simpler” Mechanisms

Now, adapt the multiagent version of the Antle and Eppen model to incorporate agent risk aversion and a nonlinear production function. Denote manager i 's utility by $u(t) - v(x, c)$, where $u'(\cdot) > 0$ and $u''(\cdot) < 0$. The manager's disutility of producing x satisfies: $v(x, c_L) < v(x, c_H)$, $v_x(\cdot) > 0$, $v_{xx}(\cdot) > 0$, and $v_x(x, c_L) < v_x(x, c_H)$. The first of these assumptions says that producing the same output level x is less costly when the cost is low than high; the second that increasing production is costly; and the third that there are decreasing returns (increasing costs) to scale. The last assumption on v ensures the

well-known single crossing property is satisfied: the constant utility curves (the isoquant curves) cross at a single point.

With agent risk aversion and the nonlinear cost function, the bankruptcy (B) constraints, output feasibility (OF) constraints, and non-negativity (NN) constraints are no longer needed. The notation is different, but this is essentially the same model studied in Demski and Sappington (1984). Demski and Sappington also assume the choice of output is decentralized to the managers so add the constraints (D): $x_{km} = x_k$ for all k, m . Program (AS-RA) formulates headquarters’ problem.

$$\underset{x,t}{Max} E[x_{km} + x_{mk} - t_{km} - t_{mk}] \quad (\text{AS-RA})$$

s.t.

$$E_{m=L,H}[u(t_{km}) - v(x_{km}, c_k)|c_k] \geq \bar{U}, \quad \forall k \quad (\text{IR})$$

$$E_{m=L,H}[u(t_{km}) - v(x_{km}, c_k)|c_k] \geq E_{m=L,H}[u(t_{nm}) - v(x_{nm}, c_k)|c_k], \\ \forall k, n \quad (\text{IC})$$

$$x_{km} = x_k, \quad \forall k, m \quad (\text{D})$$

The following proposition characterizes the solution to Program (AS-RA).

Proposition 2.3. The solution to Program (AS-RA) has the following properties:

- (i) The individual rationality constraint is binding for a high-cost manager and may or may not bind for a low-cost manager.
 - (ii) The incentive compatibility constraint is binding for a low-cost manager but does not bind for a high-cost manager.
 - (iii) $t_{LL}^* = t_{LH}^*$ and $t_{HL}^* < t_{HH}^*$.
 - (iv) $x_H^* < x_L^*$ under constant absolute risk aversion.
-

Proof. See Demski and Sappington (1984) for (i)–(iii) and Ma et al. (1988) for (iv). \square

Since it is possible for the individual rationality constraints of both types to bind (that neither type earns rents), one might wonder if the first-best solution is obtained. The first-best solution is obtained only when the managers' types are perfectly correlated. With less than perfect correlation, the payments impose risk on the managers, which is costly. The characterization of the second-best solution leads to the following corollary.

Corollary to Proposition 2.3. Under the solution to Program (AS-RA), in addition to the equilibrium headquarters intends the managers to play, there is another equilibrium in which both low- and high-cost managers produce the output intended for the high-cost type, x_H^* . From the managers' perspective, the equilibrium that has both types producing x_H^* Pareto-dominates the equilibrium headquarters intends them to play.

As Demski (2010, p. 455) writes: “[c]oordination temptations then enter, as the orchestrated competition between the managers can be turned off by playing a second and more advantageous equilibrium.” The approach to the multiple equilibrium (tacit collusion) problem proposed by Demski and Sappington (1984) is to impose dominant strategy incentives on one of the managers. With risk-averse managers, dominant strategy constraints are more costly than Bayes–Nash constraints, so headquarters' objective function is reduced.

Can headquarters design some other mechanism that eliminates the undesirable equilibrium without creating a new unwanted equilibrium and under which the equilibrium payoffs are second-best? That is, can headquarters implement the second-best solution? This question is answered by Ma et al. (1988).

Proposition 2.4. By augmenting the second-best solution of (AS-RA) with a continuum of off-equilibrium production levels, headquarters can ensure the equilibrium she intends the managers to play is unique. The equilibrium production levels and payments are as prescribed by the second-best solution.

Proof. See Ma et al. (1988). □

The following is an alternative mechanism to the one Ma et al. (1988) construct but is more typical of those commonly employed in the implementation literature. Let manager A choose $x^A \in (x_0, x_H^*) \cup \{x_H^*, x_L^*\}$.

Set $t^A = t_{km}^A$ if $x^A \in \{x_H^*, x_L^*\}$. Otherwise, choose t^A so such that:

$$u^A(t^A) = u^A(t_{HH}^{A*}) - [v^A(x_H^*, c_H) - v^A(x^A, c_H)] \\ + \Pr(c_L|c_H) \left(\frac{x^A}{x_H^*} \right) \quad \text{if } x^B = x_H^*$$

and

$$u^A(t^A) = u^A(t_{HL}^{A*}) - [v^A(x_H^*, c_H) - v^A(x^A, c_H)] \\ - \Pr(c_H|c_H) \quad \text{if } x^B = x_L^*.$$

The key idea is to offer one manager, say manager A, the option of forecasting the play of the other player, manager B. If manager B is reporting his cost is high with probability greater than $\Pr(c_H)$, manager A will find it optimal to “rat” on manager B by choosing some output $x^i < x_H^*$ when his cost is c_H .

Rajan (1992) uses a mechanism similar to the one just presented to explain cost allocation. The cost allocation mechanism gives the divisional managers a similar opportunity to rat on each other, tying them together in a way that was new to the theory of cost allocation.

The mechanism takes a short-cut. Namely, instead of ensuring manager B has a best response to manager A’s ratting that is second-best, the mechanism simply ensures manager A has no optimal ratting strategy. Any $x^A < x_H^*$ can be improved upon by one closer to x_H^* . That is, the mechanism exploits a weakness of the (Bayes) Nash equilibrium concept: best responses do not always exist when the action space is not compact. This critique applies to essentially all implementation theory as of 1992 (and most since), which relied on “integer” and “modulo” games that exploited weaknesses of solution concepts much in the same way as the above mechanism (Jackson, 1992). A reply to Jackson’s criticism is that there are certain types of real games that are not worth playing (e.g., the game of chicken as portrayed in the James Dean movie “Rebel Without a Cause”). Nevertheless, Jackson’s critique is compelling, particularly if one views implementation theory as

an attempt to understand observed institutions (instead of only what can and cannot be implemented).

Glover (1994) shows that a simpler mechanism not subject to the Jackson critique can approximately implement the second-best solution in the Demski–Sappington model. The key idea is to turn the augmented message space from “ratting” on the other manager’s lying into “confession” of one’s own lying. The confession is offered via a third output level x_M , $x_H < x_M < x_L$. The payments are chosen so that x_M is appealing to the manager as an alternative to x_H if and only if the manager’s cost is c_L . That is, x_M represents a confession to wanting to produce x_H when manager i ’s cost is c_L . Although the second-best solution is only approximately implemented (the payoffs can be made arbitrarily close to second-best), confession allows for a less demanding behavioral assumption than Bayes–Nash. Implementation can be achieved via two rounds of iteratively removing strictly dominated strategies. Each agent needs only know the other agent will not play strictly dominated strategies.

Proposition 2.5. By augmenting the second-best solution of (AS-RA) with a single off-equilibrium production level, headquarters can ensure the equilibrium she intends the managers to play is the unique strategy combination that survives two rounds of iteratively removing strictly dominated strategies. The equilibrium production levels are as prescribed by the second-best solution, and the equilibrium payments can be made arbitrarily close to second-best.

Proof. See Glover (1994). □

Arya and Glover (1996) show that the confession mechanism of Glover (1994) can be reinterpreted as participative budgeting. A budget of x_H followed by production of x_L is equivalent to production of x_M in Glover (1994). Each manager anticipates the other will budget truthfully or pad his budget and then exceed the targeted performance, which can be viewed as a confession to budget padding. The possibility of confession keeps the other manager honest.

Arya et al. (2000b) generalize the result of Glover (1994), focusing on the single crossing property which allows for exact rather than Abreu and Mastushima’s (1992) approximate (“virtual”) implementation of the first-best solution when the agents have complete information. That is, in a fairly general principal–multiagent model that includes team production, auctions, and procurement as special cases, any information that is common knowledge should be relatively easy to elicit. A key assumption is that large penalties are possible, which may be more constrained in practice. Of course, the assumption of complete information is itself questionable as a description of reality, even when it comes to specific information. Nevertheless, there are some types of accounting measurements that seem to come close (e.g., historical costs).

Arya et al. also generalize Glover’s (1994) incomplete information result to any finite number of agents and types.⁴ Arya et al. (1995) show that similar results can be obtained under a milder separability assumption that incorporates pure exchange and auctions by having the agents submit an initial message before they learn their types.⁵

2.3 Robust Mechanisms

What sort of model is a good candidate to try to develop a response to the Wilson (1987) critique discussed in the introduction? The model of Cremer–McLean seems particularly well suited. Recall that the transfers go to positive and negative infinity as the correlation goes to 0 and Cremer–McLean’s critique of their own result — that it does not seem

⁴Arya et al. (2000b) also point out that the single crossing property implies Maskin Monotonicity (a necessary condition for Nash implementation), but incentive compatibility and the single crossing property do not imply Bayesian Monotonicity (the corresponding condition for Bayes–Nash implementation). However, strict incentive compatibility and the single crossing property do imply Bayesian Monotonicity. Strict incentive constraints are also important in the recent literature on ex post implementation (e.g., Bergemann and Morris, 2008). An ex post equilibrium of a game of incomplete information is a Bayesian equilibrium that is also a Nash equilibrium of the game if the players instead had complete information (if each player knew the type of all other players).

⁵An important early paper on implementation in exchange economies is Schmeidler (1980), which showed that the set of competitive equilibria can be exactly implemented in Nash equilibrium if the agents have complete information.

reasonable that the first-best solution is so easily achieved if we are interested in explaining observed institutions.

To identify robust mechanisms, maintain a traditional expected utility framework but assume the designer knows less at the time she designs the mechanism than is typically assumed. The mechanism is designed taking expectations over the variety of possible environments (project characteristics) that might subsequently emerge. Assume the mechanism cannot later be fine-tuned. It is either impossible or prohibitively costly to design a different mechanism for each environment.

The model is the (AS-RN) model studied in Section 2.1 with additional constraints and simplifying assumptions. The most straightforward of these is: $x^A + x^B = 1$. A single unit is to be procured. This can be thought of as a multidivisional capital budgeting problem in which the manager's project proposals are mutually exclusive (ME). The earlier bankruptcy (B) constraints are again removed, which leads one to expect a result similar to Proposition 2.1 in which the first-best is achieved in dominant strategies.

It is common knowledge that c^i is equally likely to be c_L or c_H . q is the probability that $c^i = c_k$ given $c^j = c_k$. That is, q is the conditional probability the managers' costs match. Before q is known, headquarters designs a single procurement auction (capital budgeting procedure) to handle the variety of different possible correlations, knowing that q is uniformly distributed on the interval $[q', 1]$ and correlation is positive, i.e., $q' > 1/2$. The same auction is to be used for all possible correlations. This can be viewed as a headquarters that will conduct many capital budget evaluations using the same capital budgeting procedure. Designing a procedure for each situation (either by waiting until the circumstance arises or by designing a complex menu) is prohibitively costly. To make the general-purpose design meaningful, we restrict the message space of each manager to be binary. For ease of interpretation, $\{c_L, c_H\}$ is the message space.

Although the generic auction has to be designed before the correlation in the managers' environments is known, by the time a specific application arises, the correlation is common knowledge among all players. For manager i who observes c_k and q , his reporting strategy is denoted by $s^i(c_k, q)$. That is, s^i is a mapping from manager i 's private

information to his message space $\{c_L, c_H\}$. Because of the restriction on communication, the Revelation Principle cannot be applied. The auction is more decentralized than is typically assumed.⁶ We have to consider all possible reporting strategies.

Headquarters' objective is to minimize the expected transfers to the managers, $t^A + t^B$, while ensuring: the managers' reporting strategies comprise a Bayes–Nash equilibrium in their subgame (incentive compatibility), the equilibrium provides each manager with an expected utility of at least zero (individual rationality), and the probability of the managers' projects being approved, x^A and x^B , sum to 1.

As a benchmark, suppose headquarters knows q at the time of designing the auction mechanism. In this case, despite c^i being private, headquarters can extract the full surplus.

Proposition 2.6. If headquarters knows q at the time of designing the auction, a solution to Program (AS-RN) with mutually exclusive (ME) projects and without the bankruptcy (B) constraints has the following features:

(i) $x_{LL}^* = x_{HH}^* = \frac{1}{2}$, $x_{LH}^* = 1$, and $x_{HL}^* = 0$.

$$t_{HH}^* = \left(\frac{q^2(c_H - c_L)}{2(2q - 1)} \right) + \frac{c_L}{2}, \quad t_{LH}^* = t_{HH}^* + \frac{c_L}{2},$$

$$t_{HL}^* = \left(\frac{-q(1 - q)(c_H - c_L)}{2(2q - 1)} \right), \quad t_{LL}^* = t_{HL}^* + \frac{c_L}{2}.$$

- (ii) The managers earn no information rents (the payoffs are first-best).
- (iii) The incentive compatibility constraints are satisfied in dominant strategies.

Proof. See Cremer and McLean (1988). □

⁶For another model of internal auctions in which an auction is viewed as a more decentralized mechanism than a direct (revelation) mechanism, see Baiman et al. (2007) in which an auction can replicate the performance of the second-best direct revelation mechanism.

Proposition 2.6 is Proposition 2.1 adapted to procurement, or capital budgeting for mutually exclusive projects.

Return to our assumption that the auction is designed before q is known. The revised program with incentive compatibility constraints corresponding to each possible q is labeled as Program (AS-RN-Robust). Its solution is of one of two forms (Arya et al., 2009).

Proposition 2.7. The uniquely optimal solution to (AS-RN-Robust) has x^* as in Cremer–McLean. The payments are of one of two forms:

- (i) For $q' > 2/3$, t_{HH}^* and t_{HL}^* are as in Cremer–McLean with q replaced by q' and $t_{LH}^* = c_L$ and $t_{LL}^* = c_L/2$. Low-cost managers earn no rents, while high-cost managers earn rents for $q > q'$, and the incentive compatibility constraints cannot be satisfied in dominant strategies (without lowering the objective function value).
- (ii) For $q' \leq 2/3$, the auction is a (modified) second-price auction: $t_{HH}^* = c_H/2$, $t_{HL} = 0$, $t_{LH}^* = c_H/2 + c_L/2$, and $t_{LL}^* = c_L/2$. Only low-cost managers earn rents, and the incentive constraints are satisfied in dominant strategies.

Proof. See Arya et al. (2009). □

When there is little variation in the correlation the mechanism has to be robust to, the solution is similar to the known (common knowledge) correlation case. For a high-cost manager, the payments are as in Cremer–McLean, except q is replaced by q' . As a result, high-cost managers earn rents for $q > q'$. This is unusual. We are used to low-cost (high-type) managers earning rents but not high-cost managers. Here, a low-cost manager does not earn information rents. As a result of the differing payment schemes offered to low- and high-cost managers, the incentives cannot be obtained in dominant strategies without lowering headquarters' objective function value. In this sense, even a small robustness concern is important in that it converts a nonunique solution into a unique one — in this case, ruling out a dominant strategy characterization of the optimal Bayes–Nash auction.

When there is substantial variation in the correlation the mechanism has to be designed to deal with, the solution is qualitatively different. The optimal mechanism in this case is essentially a second-price (Vickrey) auction that completely ignores the correlation. The important point is that the optimal “robust” mechanism is qualitatively similar to the standard one when robustness is a minor concern but is qualitatively different when robustness is a major concern.⁷

Arya et al. (2009) explore variations. If the upper bound on the correlation is not 1, a second dominant strategy auction that exploits the correlation in the managers’ costs emerges as a middle ground.⁸ If the distribution over the correlation parameter is not uniform, the optimal reporting strategies can have an interior cutoff on q below which the managers switch from one cost report to the other for the same actual cost, although the form of the optimal auction remains the same.

So far, no attempt has been made to deal with multiple equilibria. The second-price auction does not suffer from a compelling multiple equilibrium problem, but the modified Cremer–McLean solution does. It seems that the approach of Ma et al. (1988) would be difficult to apply because of the unknown correlation. However, an approach similar to Arya et al.’s (1995) should work, since it does not rely on the correlation.

There are other approaches to the robustness problem. Reichelstein (1997) and Dutta and Reichelstein (2002) study the design of optimal performance measures in investment settings. They use robustness as a way to choose between multiple optimal performance measures. Bergemann and Morris (2005) study a general implementation model and require the mechanism to be robust to all possible higher beliefs agents might have. Bergemann and Morris (2008) study ex post implementation — Bayes–Nash equilibria of games of incomplete information that are also Nash equilibria if the agents’ information were instead

⁷The constraints that bind change at $q' = 2/3$. This is relatively straightforward in that it follows standard proof techniques. The nonstandard part of the proof arises from the earlier observation that the Revelation Principle cannot be applied.

⁸The more general specification allows the uncertainty about the correlation to be separated from its mean. Under the more restrictive specification adopted above, q' captures both.

complete. Chung and Ely (2007) study the role a maxmin objective function can play in making dominant strategy auctions optimal. They show that a dominant strategy auction that completely ignores the correlation can be the best way to guard against the worst potential beliefs.

In contrast, Arya et al. (2009) is not about robustness to higher order beliefs or other departures from the standard Bayesian approach. Robustness is instead modeled as an early design requirement, with the mechanism later to be applied to a variety of situations. The optimal robust mechanism is the one that works best on average across the variety of environments.

3

Moral Hazard

3.1 Second-Best Contracts

The problem of moral hazard is created by unobservable actions. The model is a multiagent extension of Laffont and Martimort (2001, Section 4.3). Again, a risk-neutral headquarters contracts with two risk-neutral managers, A and B . Headquarters would like to make non-negative payments (transfers) to the managers, t^A and t^B , that induce each of them to exert unobservable high effort, $e^i = H$, as a Nash equilibrium. That is, (H, H) should be preferred to (L, H) for the manager who is contemplating a deviation from H to L .

In the absence of an incentive to do otherwise, the managers would prefer to exert low effort, $e^i = L$, referred to as shirking, $L < H$. Manager i has preferences over wealth and effort that are represented by a von Neumann–Morgenstern utility function $u^i(t^i, e^i) = t^i - e^i$. The only jointly observable variables are the outputs produced by the two divisions, x^A and x^B , $x^i \in \{S, F\}$, where S denotes success and F denotes failure. Denote by t_{km}^i the payment to manager i when his division's output x^i is k and manager j 's output x^j is m ; $k, m = S, F$. There is a common shock σ . With probability σ , the outcome in both divisions will be a success, independent of effort. Effort comes into play

only with probability $(1 - \sigma)$. In particular, the probability of success is $\sigma + (1 - \sigma)p_H$ if manager i chooses high effort and $\sigma + (1 - \sigma)p_L$ if manager i chooses low effort. σ is an exogenous source of correlation in the managers' environments.

Headquarters' contracting problem is formalized in the following Linear-Program (MH).

$$\text{Min}_t E[t_{km} + t_{mk}] \quad (\text{MH})$$

s.t.

$$E[t_{km}|H, H] - H \geq \bar{U} \quad (\text{IR})$$

$$E[t_{km}|H, H] - H \geq E[t_{km}|L, H] - L \quad (\text{IC})$$

$$t_{km} \geq 0, \quad \forall k, m \quad (\text{B})$$

For simplicity, assume $\bar{U} = L = 0$. In this case, the individual rationality (IR) constraints do not bind. In addition to the (Nash) incentive compatibility (IC) constraints, the non-negative payment/bankruptcy (B) constraints are an important source of contracting frictions in this model. Headquarters makes payments to the managers, not the other way around. If negative transfers were allowed, each division could be sold to its manager for its expected value, obtaining the first-best solution.

Under the solution to Program (MH), the second-best solution, likelihood ratios are key. A single bonus is paid to each manager when the outcome with the best likelihood ratio is realized; otherwise, the manager is paid 0. This is sometimes referred to as "bang-bang" solution. A likelihood ratio captures the probability of an outcome under low effort versus its probability under high effort (e.g., p_L/p_H).

When $\sigma = 0$, the problem collapses to a twofold version of the single-manager model, and each manager is paid $H/(p_H - p_L)$ when there is a success in his division and 0 when there is a failure. This is an individual performance evaluation (IPE) contract. The expected payment is $p_H * H/(p_H - p_L) = H/(1 - p_L/p_H)$. The smaller p_L/p_H , the lower the expected cost to headquarters. The first-best is obtained if and only if $p_L = 0$.

A positive common shock ($\sigma > 0$) makes it optimal to use relative performance evaluation (RPE) in compensating the managers. If we

continued to reward each manager using IPE, the related likelihood ratio is $\frac{p_L + \sigma}{p_H + \sigma}$. If instead we use RPE, rewarding each manager only when his own outcome is a success but the other division obtains a failure, the related likelihood ratio is $\frac{(1-\sigma)p_L(1-p_H)}{(1-\sigma)p_H(1-p_H)} = p_L/p_H$. That is, RPE completely removes the noise introduced by the common shock σ .

Proposition 3.1. The solution to Program (MH) is RPE: $t_{FF}^* = t_{FS}^* = t_{SS}^* = 0$ and $t_{SF}^* = \frac{H}{(1-\sigma)[p_H(1-p_H) - p_L(1-p_H)]}$. The expected payment to each manager is $E[t^*] = \frac{H}{(1-p_L/p_H)}$.

Proof. See the appendix or Che and Yoo (2001). □

This payment scheme ensures high effort is a Nash equilibrium, but following our earlier discussion of the adverse selection model, is there another equilibrium and, in particular, another equilibrium that Pareto-dominates (from the managers' perspective) the (H, H) equilibrium headquarters intends them to play? As it turns out, the answer is no for this structure. The equilibrium headquarters wants the managers to play is essentially unique. The reason is that RPE creates a strategic substitutability in the managers' actions: each manager's increase in expected utility from changing from L to H is highest when the other manager is choosing L . The probabilities tell the story: $(p_H - p_L)(1 - p_L) > (p_H - p_L)(1 - p_H)$, i.e., H is a dominant strategy under RPE. Since H is a dominant strategy, increasing the positive payment by any arbitrarily small amount makes the equilibrium unique.

Corollary to Proposition 3.1. Under the second-best Nash solution, H is a dominant strategy.

The corollary hinges on the single-period setting. If the game were repeated, RPE presents a natural opportunity for the managers to tacitly collude on alternating between (H, L) and (L, H) , taking turns making the other look good, because of the strategic substitutability RPE creates in the managers' payoffs (see Baldenius and Glover, 2012). We will come back to this later when we consider repeated contracting.

3.2 The Multiple Equilibrium Problem and “Simpler” Mechanisms

The correlation structure in the previous section of the monograph is important. With a more general structure, it is easy to construct examples in which tacit collusion arises as a problem (Mookherjee, 1984). Consider the following numerical example in which the managers’ efforts are strategic complements under RPE: each manager’s increase in expected utility from changing from L to H is highest when the other manager is choosing H .

Actions\Outcomes		Pr(outcomes actions)			
		(F, F)	(F, S)	(S, F)	(S, S)
H, H		0.2	0.1	0.1	0.6
L, H		0.35	0.15	0	0.5
H, L		0.35	0	0.15	0.5
L, L		0.4	0.1	0.1	0.4

Suppose $H = 1$. The optimal second-best contract is $t_{FF}^* = t_{FS}^* = t_{SS}^* = 0$ and $t_{SF}^* = 1/(0.1 - 0) = 10$. The expected payment is $0.1(10) = 1$, which is first-best. Now, consider the managers’ subgame.

The Managers’ Subgame		
$e^A \setminus e^B$	H	L
H	0, 0	0.5, 0
L	0, 0.5	1, 1

(A’s Payoff, B’s Payoff)

(L, L) is also a Nash equilibrium and, from the managers’ perspective, Pareto-dominates the (H, H) equilibrium.

One potential solution is to alter the payments so that L is a dominant strategy for each manager. The optimal way of doing this is to increase the (same) bonus payment from 10 to 20. Of course, this is costly.

Ma (1988) studies Mookherjee’s (1984) multiagent model of moral hazard and shows it is possible to exactly implement the second-best solution if action combinations produce distinct probability distributions over outcomes, using another “ratting” mechanism. Arya and

Glover (1995) adapted Glover’s (1994) “confession” mechanism to deal with moral hazard.¹

Proposition 3.2. Assume $\text{Min}\{\text{Pr}(x^i = S|H, H), \text{Pr}(x^i = S|H, L)\} > \text{Max}\{\text{Pr}(x^i = S|L, H), \text{Pr}(x^i = S|L, L)\}$. Then a mechanism that has each manager forecasting his own output can approximately implement the second-best solution via two rounds of iteratively removing strictly dominated strategies.

Proof. See Arya and Glover (1995). □

Proposition 3.2’s assumption is that each division manager’s effort is more important than the other division manager’s effort in determining the given division’s output. Negative payments are not ruled out in Arya and Glover (1995), but their construction can be adapted to incorporate that additional constraint. The following is such a mechanism for the numerical example but is easily generalized under the assumption given in Proposition 3.2.

Ask each manager to forecast his own output by submitting a message $f^i \in \{S, F\}$. Make the second-best incentives strict: $t_{SF}^{*\varepsilon} = 10 + \varepsilon$ and $t_{FF}^{*\varepsilon} = t_{FS}^{*\varepsilon} = t_{SS}^{*\varepsilon} = 0$. Create any strict dominant strategy contract, say $t_{SF}^{D\varepsilon} = 20 + \varepsilon$ and $t_{FF}^{D\varepsilon} = t_{FS}^{D\varepsilon} = t_{SS}^{D\varepsilon} = 0$. Choose any $\varepsilon > 0$ arbitrarily small. Manager i ’s pay is as follows:

$$\begin{aligned} &\text{If } f^j = S \text{ and } f^i = S: t^i = t^{*\varepsilon} + \varepsilon \text{ if } x^i = S \text{ and } t^i = t^{*\varepsilon} \text{ if } x^i = F. \\ &\text{If } f^j = S \text{ and } f^i = F: t^i = t^{*\varepsilon} \text{ if } x^i = S \text{ and } t^i = t^{*\varepsilon} + 1.2\varepsilon \text{ if } \\ &\quad x^i = F. \\ &\text{If } f^j = F \text{ and } f^i = S: t^i = t^{D\varepsilon} + \varepsilon \text{ if } x^i = S \text{ and } t^i = t^{D\varepsilon} \text{ if } \\ &\quad x^i = F. \\ &\text{If } f^j = F \text{ and } f^i = F: t^i = t^{D\varepsilon} \text{ if } x^i = S \text{ and } t^i = t^{D\varepsilon} + 1.2\varepsilon \text{ if } \\ &\quad x^i = F. \end{aligned}$$

Under this mechanism, the managers’ subgame is the following.

¹Arya et al. (1997b) study team production, exploiting the known reservation utility associated with rejecting the employment contract. The (IR) constraint both constrains (as one might expect) and eases (as one might not expect) the implementation problem because of the known utility associated with contract rejection. See also Jackson and Palfrey (2001).

The Managers' Subgame

$(e^A, f^A) \setminus (e^B, f^B)$	(H, F)	(H, S)	(L, F)	(L, S)
H, F	$1 + 0.1\varepsilon + 0.36\varepsilon$	$0.1\varepsilon + 0.36\varepsilon$	$2 + 0.15\varepsilon + 0.42\varepsilon$	$0.5 + 0.15\varepsilon + 0.42\varepsilon$
H, S	$1 + 0.1\varepsilon + 0.7\varepsilon$	$0.1\varepsilon + 0.7\varepsilon$	$2 + 0.15\varepsilon + 0.65\varepsilon$	$0.5 + 0.15\varepsilon + 0.65\varepsilon$
L, F	0.6ε	0.6ε	$2 + 0.1\varepsilon + 0.6\varepsilon$	$1 + 0.1\varepsilon + 0.6\varepsilon$
L, S	0.5ε	0.5ε	$2 + 0.1\varepsilon + 0.5\varepsilon$	$1 + 0.1\varepsilon + 0.5\varepsilon$

Manager A's Payoffs (B's are Symmetric)

In the above subgame, (H, S) strictly dominates (H, F) and (L, F) strictly dominates (L, S) . Each manager has strict dominant strategy incentives to forecast his own division's outcome honestly. Equivalently, each manager has strict dominant strategy incentives to confess to having shirked if and only if he has. After these strictly dominated strategies are eliminated, the managers' subgame is as follows.

The Managers' Subgame
after Deleting Strictly Dominated Strategies

$(e^A, f^A) \setminus (e^B, f^B)$	(H, S)	(L, F)
H, S	$0.1\varepsilon + 0.7\varepsilon$	$2 + 0.15\varepsilon + 0.65\varepsilon$
L, F	0.6ε	$2 + 0.1\varepsilon + 0.6\varepsilon$

Manager A's Payoffs (B's are Symmetric)

In the above game, (H, S) strictly dominates (L, F) , i.e., working strictly dominates shirking. By choosing $\varepsilon > 0$ as small as headquarters desires, the equilibrium payments can be made arbitrarily close to the second-best payments. The second-best solution is approximately implemented via two rounds of iteratively removing strictly dominated strategies.

3.3 Optimal Use of Additional Performance Measures

As noted in the previous subsection of the monograph, one approach to the multiple equilibrium problem is to turn up the power of incentives in order to ensure the desired behavior is a dominant strategy. Despite the early work of Demski and Sappington (1984) on adverse selection, the dominant strategy approach to preventing tacit collusion is relatively under explored in principal-agent models. The dominant strategy approach has the performance measures taking on a qualitatively new role, which is best illustrated with uncorrelated individual performance measures.

In addition to x^A and x^B , which are allowed to have a general structure (as in the numerical example in the previous subsection of the paper), introduce individual performance measure y^A and y^B ; $y^A, y^B = L, H$. Suppose $\Pr(y^A = H|e^A = k) = q_k^A$ and $\Pr(y^B = H|e^B = m) = q_m^B$. Since these are individual performance measures, y^i is a function of e^i but not e^j ; $i \neq j$. (If the only performance measures are y^A and y^B , then the problem reduces to a twofold repetition of the principal–single agent contracting problem, and tacit collusion is not an issue.)

In order to make H a dominant strategy for each manager, it is optimal to condition each manager’s pay on this new information in a way that produces higher-powered incentives when the new information indicates the other manager is choosing L rather than H . Conditioning manager i ’s payment positively on y^i and negatively on y^j allows the power of incentives to be turned up selectively — when the managers are playing (L, L) but not when they are playing (H, H) . Although the other manager’s individual performance measure is not informative of the given manager’s effort in the sense of Holmstrom (1979), the possibility of collusion creates an endogenous correlation that makes such conditioning optimal.

Proposition 3.3. Suppose (L, L) is an equilibrium that Pareto-dominates the (H, H) equilibrium in the managers’ subgame under the second-best (Nash) solution. Then the optimal dominant strategy contract has manager i ’s pay depending positively on y^i and negatively on y^j ; $i, j = A, B$; $i \neq j$.

Proof. See the appendix. □

3.4 Bailouts and Coordinated Moral Hazard

Return to the earlier model in which only x^A and x^B are available for contracting and are correlated only through the common shock σ . Assume x^A and x^B are individual performance measures, so $\sigma = 0$. Headquarters receives an early (and, for simplicity, perfect) read on the outputs of the divisions. The early read is not verifiable and hence cannot be contracted on. Headquarters is able to bail the managers

out and guarantee an outcome of (S, S) , success in both divisions, at a personal cost of $C > 0$. Suppose the size of C is such that headquarters will bail the managers out if and only if the early read is (F, F) . This intervention is also not verifiable and, hence, noncontractible. The contract offered to the managers depends only on their final divisional outcomes.

In the previous section of the monograph, when $\sigma = 0$, an optimal contract was individual performance evaluation (IPE): $t_{FF} = t_{FS} = t_F = 0$, $t_{SF} = t_{SS} = t_S = \frac{H}{p_H - p_L}$. Because of headquarters' intervention, this contract is no longer incentive compatible. The bonus must be increased to $t_S^{IPE} = \frac{H}{p_H(p_H - p_L)}$.

Consider an example with $H = 1$, $p_H = 0.8$, and $p_L = 0.6$. Without the bailout option, IPE would be an optimal contract with $t_S = 5$. With the bailout option, the second-best IPE contract requires t_S to be increased to 6.25. The managers' subgame is as follows, taking headquarters' bailout decision as given.

The Managers' Subgame

$e^A \backslash e^B$	H	L
H	4.25, 4.25	4.5, 4.25
L	4.25, 4.5	4.75, 4.75

(A's Payoff, B's Payoff)

Headquarters' bailout creates an endogenous correlation in the managers' environments and a multiple equilibrium problem. Moreover, the (L, L) equilibrium Pareto-dominates the (H, H) equilibrium in the managers' subgame. The optimal solution is RPE, since it ensures the managers are rewarded only when they are not bailed out. The optimal contract depends both qualitatively and quantitatively on whether or not the bailout option is present.

Proposition 3.4. In the (MH) model with uncorrelated production technologies ($\sigma = 0$) and headquarters' bailout of all potential (F, F) outcomes:

- (i) The optimal Nash incentive compatible IPE contract is

$$t_F^{IPE} = 0, \quad t_S^{IPE} = \frac{H}{p_H(p_H - p_L)}.$$

- (ii) Under t^{IPE} , (L, L) is also a Nash equilibrium and Pareto-dominates the (H, H) equilibrium in the managers' subgame.
- (iii) The optimal solution is RPE.

Proof. See the appendix. □

The problem can be more difficult to solve in other settings. Arya and Glover (2006) consider a joint production setting in which headquarters bails the managers out only for the worst of three possible outcomes: $F < I$ (intermediate) $< S$. Consider the following numerical example.

		Pr(outcomes actions)		
		F	I	S
Actions\Outcomes	H, H	0.2	0.4	0.4
	L, H	0.3	0.5	0.2
	H, L	0.3	0.5	0.2
	L, L	0.7	0.3	0

Without headquarters' bailout, the optimal contract is to set $t_S = H/(0.4 - 0.2) = 5H$ with an expected payment per manager of $2H$. Under the promise of a potential bailout, with headquarters' bailout just before F is observed (converting F into S), t_S must be increased to $t_S = H/(0.6 - 0.5) = 10H$ to ensure that (H, H) is Nash equilibrium. However, there is now an (L, L) equilibrium that Pareto-dominates the (H, H) equilibrium. Given headquarters' intervention strategy, the probability of S is maximized when the managers choose (L, L) , so there is no hope of using a bonus based on the S outcome to induce the managers to play (H, H) as a unique equilibrium. The best headquarters can hope to do is to motivate (L, H) or (H, L) by paying the managers a bonus when I is realized, unless she finds some way to tie her own hands and not intervene.

Under joint production, a potential solution is to install a coarse information system, so the situation cannot be monitored so closely (Arya and Glover, 2006). Coarse information can make it in headquarters' best interest to limit her intervention in the firm's operations.

Realizing they will not be bailed out, the managers no longer find it worthwhile to collude.

There has been a push (including regulation and stock exchange requirements) in recent years to increase the monitoring role of corporate boards, primarily by increasing the role of independent directors. Although the use of independent directors as monitors has appeal, the possibility of intervention by the board can also provide the management team with opportunities to collude. There are existing models of corporate governance that focus on the role excessive oversight by independent directors can have in limiting a CEO's benefit of producing and/or communicating information to an unfriendly (overly independent) board (e.g., Adams and Ferreira, 2007). As far as I know, the impact of board oversight on collusion among top executives has not yet been studied.

Proposition 3.4 suggests that increased board oversight might lead to an increased role for relative performance evaluation of top executives of the same firm (including, perhaps, increased CEO turnover and internal promotions to the CEO position), but not because a more active board facilitates such evaluations. Instead, a board's increased intervention might create an opportunity for coordinated moral hazard among top executives unless the increased intervention is accompanied by increased relative performance evaluation. The broader point is that extending models of board monitoring (and advising) to include a management team (rather than a single manager) may lead to new insights and empirical predictions.

3.5 Robust Mechanisms

Turning to robustness under moral hazard, suppose that two principals are each writing a contract with one of two agents. Because of the common shock, σ , in the agents' environments, the principals would each like to use relative performance evaluation: $t_{FF}^{RPE} = t_{FS}^{RPE} = t_{SS}^{RPE} = 0$, $t_{SF}^{RPE} = \frac{H}{(1-\sigma)[p_H(1-p_H)-p_L(1-p_H)]}$, with an expected payment to each agent of $E[t^{RPE}] = \frac{H}{(1-p_L/p_H)}$, completely removing the noise introduced by σ . However, neither principal knows the other agent's (other firm's) productivity. Principal i knows only that p_H^j is uniformly

distributed on $[\bar{p}_H - \varepsilon, \bar{p}_H + \varepsilon]$. The agents, being closer to the productive environment, know each other's productivity exactly. All other parameters, including σ , are common knowledge. So, in the RPE contract, the $1 - p_H$ terms are replaced with $1 - \bar{p}_H - \varepsilon$ terms, which increases the expected bonus. If ε is large enough, then the principals prefer individual performance evaluation to relative performance evaluation.

Proposition 3.5. If ε is large enough ($\varepsilon > \frac{(1-p_H^B)\sigma}{p_H^A + \sigma(1-p_H^A)}$), principal A prefers individual performance evaluation to relative performance evaluation.

Proof. See the appendix. □

The evidence on relative performance evaluation at the executive level is mixed (e.g., Antle and Smith, 1986; Janakiraman et al., 1992). One explanation is that positive weights on both a firm's and its competitors' performance can be optimal as a way to soften competition (Aggarwal and Samwick, 1999). A lack of robustness is another potential explanation: those designing incentives may have poorer information than executives about the industry and/or other details that make using relative performance evaluation challenging.²

²As an aside, we typically treat correlation as exogenous, while it is endogenous in practice. Managers can use both operating and financial instrument choices to change the correlation once the incentive contract is in place.

4

Implicit Contracts

4.1 Two-Period Relational Contracts Between Agents

Groups of workers often have much better information about their individual contributions than the employer is able to gather. Group incentives then motivate the employees to monitor one another and to encourage effort provision or other appropriate behavior... the possibility of withholding help from slackers can be very effective in providing incentives for members of the group to adhere to the group norms (Milgrom and Roberts, 1992, p. 416).

Continue with our earlier model of moral hazard, except the managers are now engaged in team production and repeat the relationship over a two-period horizon. They produce a single output x that can take a value of success (S) or failure (F). Denote $\Pr(S|LL)$ by p_L , $\Pr(S|LH) = \Pr(S|HL)$ by p , and $\Pr(S|HH)$ by p_H .

The key idea is to focus on the role of mutual monitoring and implicit (self-enforcing) contracting the managers do with each other — to turn the tables on multiple equilibria so that they help rather than hurt headquarters. For now, ignore discounting to keep things simple.

The explicit contract headquarters offers the managers can be different in each period, $t_1 = (t_{1F}, t_{1S})$ in the first period and $t_2 = (t_{2F}, t_{2S})$ in the second period, but is short-term in the sense that it depends only on that period's output. The question is, can headquarters design contracts that induce the managers to engage in mutual monitoring (cooperation) in the first period, which benefits headquarters by reducing expected compensation in the first period, by creating multiple equilibria in their second-period subgame?

In the second period, there is no future play the managers can use to enforce cooperation, so headquarters must satisfy the optimal single-period Nash incentive compatible constraints. The least costly such contract is: $t_2 = t^{\text{Individual}}$, where $t_F^{\text{Individual}} = 0$ and $t_S^{\text{Individual}} = \frac{H}{p_H - p} \cdot t^{\text{Individual}}$ happens to create the needed multiple equilibria in the second period. The form of the multiple equilibria in the second-period game depends on whether the managers' actions are strategic complements ($p_H - p > p - p_L$) or strategic substitutes ($p_H - p < p - p_L$). An example of a team production setting with a strategic complementarity is an interdisciplinary team in which each team member's effort is essential. An example of team production setting with a strategic substitutability is a project that inherently requires multiple team members but in which effort is interchangeable and there are overall decreasing returns to effort.

To see what these games look like, suppose $H = 1$, $p_L = 0.3$, and $p_H = 0.8$. First, consider the case of a strategic complementarity, say $p = 0.4$. In this case, $t_S^{\text{Individual}} = 2.5$, and the managers' second-period subgame is the following.

The Managers' Period 2 Subgame
under Strategic Complements

$e^A \backslash e^B$	H	L
H	1, 1	0, 1
L	1, 0	0.75, 0.75

(A's Payoff, B's Payoff)

In addition to the (H, H) equilibrium, (L, L) is also an equilibrium. Now, consider the case of strategic substitutes, say $p = 0.7$. In this case, $t_S^{\text{Individual}} = 10$, and the managers' second-period subgame is the following.

The Managers' Period 2 Subgame
under Strategic Substitutes

$e^A \backslash e^B$	H	L
H	7, 7	6, 7
L	7, 6	3, 3

(A's Payoff, B's Payoff)

In the case of strategic substitutes, in addition to the (H, H) equilibrium, (L, H) and (H, L) are also equilibria.

In both cases, there are multiple equilibria, and each manager may be willing to punish the other manager for bad behavior in the first period. The punishment is perhaps more suspect in the complements case, since it requires the managers to punish themselves along with the other manager (Bernheim and Whinston, 1998). If we count on the managers playing the Pareto-optimal equilibrium in their overall subgame, why would they play differently in their second-period subgame? In defense of such self punishments, when players punish by reverting to the stage game (one-shot game) equilibrium in infinitely repeated games, such play also sometimes has players punishing themselves along with the other players (e.g., in the infinitely repeated Prisoner's Dilemma). The approach is to focus on what can and cannot be achieved as a subgame perfect equilibrium without requiring consistency in the criteria the players use to select among equilibria at different points in time. Although beyond the scope the analysis here, it seems intuitive to think that real players might adopt different criteria before another player's deviation from a tacit agreement than they do after such a deviation.

Suppose headquarters offers group (cooperative) incentives, $t^{\text{Cooperative}}$, in the first period, ensuring only that the managers each receive a higher payoff from (H, H) than from (L, L) : $t_F^{\text{Cooperative}} = 0$ and $t_S^{\text{Cooperative}} = \frac{H}{p_H - p_L}$. For our example, $t_S^{\text{Cooperative}} = 2$. The following are the managers' payoffs from period one play, first under strategic complements and then under strategic substitutes.

The Managers' Period 1 Payoffs
under Strategic Complements

$e^A \backslash e^B$	H	L
H	0.6, 0.6	-0.2, 0.8
L	0.8, -0.2	0.6, 0.6

(A's Payoff, B's Payoff)

The Managers' Period 1 Payoffs
under Strategic Substitutes

$e^A \backslash e^B$	H	L
H	0.6, 0.6	0.4, 1.4
L	1.4, 0.4	0.6, 0.6

(A's Payoff, B's Payoff)

Under both strategic complements and strategic substitutes, each manager is tempted to free-ride — to choose low effort when his team member is choosing high effort. However, if individual incentives are used in the second period, the resulting multiple equilibria provide an opportunity for “tit-for-tat” retaliation that always provides a sufficiently large punishment that (H, H) can be sustained in the first period. That is, H in the first period is a best response if the other manager will reciprocate your H in the first period with H in the second period but will punish your L in the first period with his own L in the second period.

Headquarters finds fostering multiple equilibria optimal. In fact, the incentive scheme just outlined is optimal among all possible subgame perfect ones: $(t_1^*, t_2^*) = (t^{\text{Cooperative}}, t^{\text{Individual}})$. Individual incentives in the second period always provide a punishment that is more than large enough to enforce group incentives in the first period, since individual incentives are more high-powered.

Proposition 4.1. In the two-period repetition of the (MH) model with a team performance measure, the optimal contract is $(t_1^*, t_2^*) = (t^{\text{Cooperative}}, t^{\text{Individual}})$.

Proof. See Arya et al. (1997a). □

Benoit and Krishna (1985) study the finitely repeated play of exogenous games in which multiple equilibria allow the players to sustain behavior in earlier periods that is not equilibrium play in the one-shot game. Arya et al. (1997a) endogenize this view of finite repetition, allowing the optimal contract to place the managers in different games in each of the periods.

Gibbons and Murphy (1992) document that “explicit contractual incentives are high when implicit career concerns are low, and vice

versa.” Explicit incentives are strongest for managers near retirement. Their evidence is consistent with Proposition 4.1. A related experimental study is Towry (2003), which examines the effectiveness of social ties on mutual monitoring relationships, using a two-period incentive scheme with essentially the same properties as our strategic complements setting to motivate mutual monitoring.

What happens if individual performance measures are available for contracting? If the managers’ environments are uncorrelated ($\sigma = 0$), JPE is just as efficient as RPE (or IPE), so we can use JPE in the same way it was used in the team production setting to induce mutual monitoring. When the managers’ environments are correlated ($\sigma > 0$), the common shock increases expected payments under JPE but not under RPE, since RPE filters the common shock out. As a result, there is a cutoff on the correlation above which individual incentives and RPE dominate cooperative incentives and JPE. The following proposition is derived by solving for the correlation that equates the expected payment under JPE (with cooperation) and RPE (with individual incentives).

Proposition 4.2. In the two-period repetition of (MH) model with individual performance measures:

- (i) If $\sigma \leq \frac{p_H p_L}{2 + p_L + p_H p_L}$, the optimal contract uses only JPE in the first period and a mix of JPE and IPE in the second period in order to induce the managers to mutually monitor each other. In the second period, the optimal weight on JPE is $\frac{p_H}{p_H + p_L}$ and on IPE is $1 - \frac{p_H}{p_H + p_L}$.
- (ii) If $\sigma > \frac{p_H p_L}{2 + p_L + p_H p_L}$, the optimal contract uses only RPE in both periods and forgoes the managers mutually monitoring each other.

Proof. See the appendix. □

Consider the following numerical example: $p_L = 0.8$, $p_H = 0.9$, and $H = 1$. If $\sigma = 0.1$, mutual monitoring is optimal. In the first period, only JPE is used: $t_{SS}^{\text{Cooperative}} = \frac{H}{(1-\sigma)(p_H p_H - p_L p_L)} = 6.54$. In the second

period, a mix of JPE and IPE is used. The second-period JPE contract is $t_{SS}^{\text{Individual}} = \frac{H}{(1-\sigma)(p_H-p_L)p_H} = 12.35$, since the second-period incentives have to motivate second-period effort. The second-period IPE contract is $t_S^{\text{Individual}} = \frac{H}{(1-\sigma)(p_H-p_L)} = 11.11$. The optimal weight on the second-period JPE contract is $p_H/(p_H + p_L) = 0.53$, where $p_H/(p_H + p_L)$ is derived by equating the first-period benefit of free riding to the second-period punishment that can be imposed in response. The expected payment to each manager over the two-period relationship is 15.59. If instead, RPE is used in each period (forgoing the benefit of mutual monitoring), the expected payment to each manager is $2/(1 - p_L/p_H) = 18$.

If $\sigma = 0.3$, RPE is optimal. Recalculate the numbers under mutual monitoring to check: the first-period JPE contract is $t_{SS}^{\text{Cooperative}} = 8.40$, the second-period JPE contract is $t_{SS}^{\text{Individual}} = 15.87$, the second-period IPE contract is $t_S^{\text{Individual}} = 14.23$, the optimal weight on the second-period JPE contract is 0.53 (as before), and the expected payment to each manager over the two-period relationship is 20.82. RPE again results in an expected payment of 18, since it filters out the effect of σ .

In Arya et al. (1997a), the managers are risk averse and operate either a team production or uncorrelated individual technologies. In their team production setting, essentially the same result as Proposition 4.1 emerges. Proposition 4.2 is new.

Itoh (1993) provides a similar result to Proposition 4.2 in a single-period setting under the assumption that the agents observe each other's actions and can write explicit side contracts with each other on their actions. In Itoh (1993), the agents' side contracting helps the principal. In contrast, in models of hidden information with explicit side contracting among agents, side contracting typically hurts the principal (e.g., Tirole, 1986). Itoh (1993) and Tirole (1986) can be viewed as reduced form models intended to capture the implicit incentives that would arise from repeated play. Arya et al. (1997a) model that repeated play.

4.2 Infinite Horizon Relational Contracts Between Agents

Now, turn to an infinite repetition of the relationship. All parties share the same discount rate r , or discount factor $d = 1/(1 + r)$. The

optimal contract depends critically on whether the managers' actions are strategic complements or substitutes. The reason is that strategic substitutes can make it optimal for the managers to tacitly collude on alternating between (H, L) and (L, H) , taking turns working while the other manager shirks.

Proposition 4.3. In the infinitely repeated (MH) model with a team performance measure:

- (i) If the managers' actions are strategic complements,

$$t_S^* = \frac{H(1+r)}{(p_H-p_L)+r(p_H-p)}.$$
 - (ii) If the managers' actions are strategic substitutes,

$$t_S^* = \text{Max} \left\{ \frac{H(1+r)}{(p_H-p_L)+r(p_H-p)}, \frac{H}{(2+r)(p_H-p)} \right\}.$$
 - (iii) As $r \rightarrow 0$, $t_S^* \rightarrow \frac{H}{(p_H-p_L)}$ under strategic complements.
 - (iv) As $r \rightarrow 0$, $t_S^* \rightarrow \frac{H}{2(p_H-p)} > \frac{H}{(p_H-p_L)}$ under strategic substitutes.
-

Proof. See the appendix. □

As far as I know, Proposition 4.3 is also new and reaches a qualitatively different conclusion from the two-period model of Arya et al. (1997a). In particular, actions that are strategic substitutes create a collusion problem that does not arise in the two-period model and limits the benefits that can be achieved from mutual monitoring.

Stepping away from the model, extra care is needed in designing incentives for team settings in which there is a strategic substitutability. For example, in assigning group projects, a professor may worry students will take turns doing the work rather than each team member working. The professor might require a group presentation in addition to a paper as a way of converting a task with a strategic substitutability into one with a strategic complementarity.

Another possible extension is to study the impact of information on the collusion opportunities. For example, suppose the managers

are undertaking a task that, on average, exhibits a strategic complementarity. A public information system that reveals that the managers' actions are complements or substitutes may have negative value because it creates new opportunities for collusion. In general, the impact of information (public or private) on implicit collusion seems to be an unexplored topic.¹

Che and Yoo (2001) study the individual production setting in an infinite horizon setting. Their main result, stated here as Proposition 4.4, is similar to Proposition 4.2. The incremental contribution of Proposition 4.2 comes from its two-period model, which allows it to capture changes in contracts over time that cannot be obtained from the infinitely repeated model (with a stationary view).

Proposition 4.4. In the infinitely repeated (MH) model with individual performance measures, there is a $r(\sigma)$ such that: JPE is optimal for $r < r(\sigma)$.

Proof. See Che and Yoo (2001). □

Giving up on mutual monitoring, RPE is the natural alternative to JPE. However, RPE creates a strategic substitutability in the payoffs. In terms of the effect on collusion, a payoff strategic substitutability has the same effect as a productive substitutability in the team output setting — the managers are tempted to collude on the alternating equilibrium. As a result, there is also a region in which IPE is optimal. IPE does not eliminate the effect of the common shock nor does IPE motivate mutual monitoring, but IPE's immunity to collusion can make it preferred to RPE. Roughly stated, the possibility of tacit collusion can be viewed as providing another foundation for dominant strategy incentives, viewing IPE as a special case of dominant strategy incentives.

¹There are many studies in accounting and economics on the value of public or private information, but they have a different focus. One recurring theme in accounting is that information system design can be valuable as a commitment device (e.g., Baiman, 1975; Arya et al., 1997c, 1998; Arya and Glover, 2006). Although collusion is not an issue in Arya et al. (1997c), the strategic complementarity/substitutability between the principal and agent's actions is a key determinant of the value of information in their model.

4.3 Relational Contracts with and Between Agents

So far, the principal and agents have had very different roles. The principal offers the agents explicit contracts that set the stage for their interactions, including implicit contracting. In practice, the principal also offers implicit contracts (promises) to agents, and these promises must also be credible in order to have value (Bull, 1987; Baker et al., 1994; Levin, 2003; MacLeod, 2003). Discretion is common in bonus awards (Murphy and Oyer, 2003). One way implicit promises are incorporated into incentives is in using subjective (non-verifiable) performance measures to reward managers. Such subjectively determined rewards appear to be used, at least in part, to foster interactions between managers that are not captured by traditional objective (verifiable) performance measures (Gibbs et al., 2004).

One approach to making such implicit promises credible is a bonus pool. Under a bonus pool, headquarters makes the size of the pool contingent on verifiable information such as firm-wide earnings and then doles out bonuses from the pool using non-verifiable information. Since the total amount of the bonus paid is independent of the non-verifiable information, headquarters' promise to divide the bonus pool as she promised is credible. For the most part, the bonus pool literature has focused on single-period settings (e.g., Baiman and Rajan, 1995; Rajan and Reichelstein, 2006, 2009; Ederhof et al., 2010).

Baldenius and Glover (2012) study dynamic bonus pools, focusing on the opportunities a strategic complementarity or substitutability in the verifiable team output creates for managers to collude. In addition to the verifiable team output, headquarters also directly observes each manager's actions, but this observation is non-verifiable. As before, when the managers' actions are strategic substitutes in the verifiable team output, the most difficult collusion among the managers to break up is the alternating equilibrium: they take turns working with the working manager receiving the entire bonus pool. The informativeness of the team performance measure helps, but preventing collusion is still costly.

When the managers' actions are strategic complements in the team output, the pressing collusion has them both shirking: they split the

bonus pool while avoiding the cost of effort. Despite the seeming advantage to conditioning the size of the bonus pool on the verifiable team output, it can be optimal not to. The reason is that a large strategic complementarity makes it extremely costly to motivate a manager to unilaterally deviate from the (L, L) equilibrium. The same is true about the alternating equilibrium under strategic substitutes: it is costly to move one manager away from L when the other manager is choosing H . However, the likelihood ratio that comes into play to break up the alternating collusion under strategic substitutes is the same likelihood ratio that comes into play on the equilibrium path. For strategic complements, the probabilities that come into play to break up the (L, L) equilibrium are different than are in play on the equilibrium path. So, conditioning the size of the bonus pool on the verifiable output is always optimal under strategic substitutes but is not optimal under a large strategic complementarity. The broader lesson is that we have to evaluate the information quality of a control system both on- and off-the-equilibrium path.

Glover and Xue (2012) search for optimal contracts rather than the exogenous bonus pools of Baldenius and Glover (2012). The model is the infinitely repeated model of moral hazard with individual performance measures (in which the managers observe each other's actions) studied in this monograph, except that all contracts — even those offered by headquarters — are implicit. There is no verifiable performance measure. The only means the managers have of retaliating against a headquarters that does not honor her promise to reward them in a given period is to choose low effort in all future periods. When the discount rate is small, headquarters' incentive compatibility constraints do not bind, and she can pay the managers as she would under full commitment (JPE as in Proposition 4.4, since $\sigma = 0$). At the other extreme of an arbitrarily large discount rate, headquarters' promises are not credible unless she adopts a bonus pool payout.

The more interesting case is an intermediate discount rate. For an intermediate discount rate, headquarters may choose to reward poor performance (F, F) even when her own incentive compatibility constraint does not require it. Rewarding joint failure can be desirable because it makes the managers' payoffs strategically independent.

With strategic independence, all collusion problems — joint shirking (L, L) and alternating between (L, H) and (H, L) — are equally costly to deal with. Since headquarters has to deal with the mostly costly of the possible collusion arrangements, making them all equally costly to prevent can be optimal.

With uncorrelated production technologies ($\sigma = 0$), strategic independence is ensured by setting $t_{SS} - t_{FS} = t_{SF} - t_{FF}$. Without the concern of collusion among the managers, the optimal way to relax headquarters' incentive constraint is to set $t_{SF} = 2t_{SS}$, i.e., to use a mix of RPE and IPE. However, the RPE component would make the managers' actions strategic substitutes and, hence, the alternating collusion is more costly to eliminate. It can be less costly to place a heavier weight on JPE, even though it comes with the cost of also increasing the t_{FF} payment in order to maintain headquarters' incentive compatibility constraint. Creating strategic independence (with IPE as a special case) is a way of dealing with collusion.

The possible optimality of not conditioning the size of the bonus pool on the verifiable team output in Baldenius and Glover (2012) and the positive payment for a failure in both divisions before headquarters' incentive compatibility constraints require it in Glover and Xue (2012) can be seen as providing explanations for pay without performance (e.g., Bebchuk and Fried, 2006).

5

Conclusion

I conclude with suggestions for future research, focusing mostly on extensions of the ideas to financial reporting and financial reporting regulation in particular, although there are also some suggestions for additional research on managerial accounting. The discussion is organized around the three main themes of the monograph.

As a precursor, suppose a financial reporting regulator such as the U.S. Securities and Exchange Commission (SEC) would like to encourage a communication rather than a compliance culture (norm) among financial statement preparers. Many SEC enforcement actions can be described as ratting mechanisms (similar to the complete information results in Ma, 1988; Arya et al., 2000b). Disgruntled former employees provide evidence their former employer adopted bad reporting behavior.

Other SEC enforcement actions have been criticized as “wild-cattling,” which is a term borrowed from the oil and gas industry to refer to situations in which drillers drill for oil where there is no reason to expect there to be any. The use of the term is somewhat misplaced: the SEC’s investigation of an entire industry is typically triggered by a first discovery of bad reporting behavior by one or more firms in

the industry. Building on Proposition 3.3, it seems likely to be optimal to use an investigation into one firm that produces evidence of bad reporting behavior as a trigger for investigating other firms when the investigations are designed to deter collusion on a bad reporting norm.

5.1 “Simpler” Mechanisms

It seems unlikely the modern firm, which is characterized by a separation of ownership and management, would be viable without auditor-verified financial statements. Auditor independence is central to the philosophy of auditing (Mautz and Sharaf, 1961). Antle (1984) articulated various economic definitions of auditor independence. In Arya and Glover (1997), a separability condition again facilitates the use of a particularly simple mechanism in maintaining auditor independence. Antle’s auditing model includes public information beyond the information produced by the accounting system. Arya and Glover add the assumption that, for each agent, there is at least one component of the nonaccounting information that provides direct information only about him. What can be accomplished by simple mechanisms under weaker conditions?

Arya and Glover’s mechanism has the manager and/or auditor confessing in the courts to their own bad behavior rather than their conjectures about the other’s bad behavior (ratting). Testifying about one’s own actions rather than conjectures about other’s behavior is consistent with the Opinion Rule, which requires testimony be limited to facts rather than conjectures (except the opinions of experts). Does the Opinion Rule have a broader foundation in mechanism design, for example, in allowing for less demanding behavioral assumptions than Nash?

In tax regulation, amnesty programs allow for a milder punishment (if any) when a taxpayer voluntarily comes forward to pay delinquent taxes. At the SEC, the Office of the Chief Accountant’s preclearance process for emerging accounting issues can be viewed as a confession and amnesty program. Are these and other similar programs used to deter collusion?

One obvious consideration that emerges from thinking about applying this monograph’s ideas to real-world settings is that most of the results are from single-period models, while real-world settings often involve repeated interactions. Can ratting and confession mechanisms be adapted to deal with repeated interactions?¹

5.2 Robust Mechanisms

A ‘hard’ measure is one constructed in such a way that it is difficult for people to disagree . . . In order to limit the room for dispute, three ingredients are necessary. One is that the measurement process must begin with verifiable facts. To base a measurement on fictions, hypotheses, opinion or unverifiable facts invites disagreement. Second, the measurement process must be well-specified to enable the parties to judge unambiguously which measurement rules for transforming facts into figures are justifiable and which are not. The verifiability of input to the measurement process does not contribute to hardening the measure if the method of producing figures from input is arbitrary. Third, the number of justifiable rules should be restricted (Ijiri, 1975, p. 36).

George O. May chaired the committee that coined the term “accepted accounting principles” (AICPA, 1934). May later wrote that

¹The *Oxford English Dictionary* (first edition) defines ratting as “[d]esertion of one’s party or principles.” In one of his letters to his brother John, Thomas Carlyle complained that a literary critic called him “the supremest German scholar in the British Empire” and that one of his essays had been described as a “splendid instance of literary ratting.” Perhaps, Carlyle did not see himself as a deserter of English literature, despite his affinity for German literature and Goethe in particular. (Hearn, 1924 writes that Carlyle’s *Sartor Resartus* seems to have been inspired by a single stanza from Goethe’s *Faust*.) It seems natural to associate ratting with the end of a relationship (e.g., whistle-blowing by former or soon-to-be former employees). If the play is repeated, the agents may tacitly collude to suppress the ratting. The confession mechanisms do not require that confession by one or more agents triggers punishments for other agents. Instead, an agent’s confession triggers dominant strategy incentives for the other agents, and the dominant strategy incentives may increase payoffs. Nevertheless, it is unclear how to adapt the confession mechanisms to repeated play.

“the plan called for a fairly full disclosure of the methods of accounting adopted by each listed company” and lamented that it “is widely used today in cases in which it is not appropriate because there is no source from which even a general sense of the principles can be obtained” (May, 1958). May’s concern seems to be about what Ijiri later called hardness (a lack thereof) and information asymmetries about hardness between financial statement preparers and users. The Accounting Principles Board was created as a partial remedy to these hardness problems. As fair value remeasurements and accounting estimates have become increasingly important, dealing with information asymmetries about hardness has become even more challenging.

What can robust mechanisms tell us about the consequences of large information asymmetries about the hardness of accounting measurements? Glover (2004) and Glover et al. (2005) were intended as a start. Glover et al.’s main result is that, when there is a large information asymmetry about the hardness of one of two measurements, the optimal contract may treat that measure in a qualitatively different manner than when the information asymmetry is small.

Generally Accepted Accounting Principles (GAAP) are themselves thought to be robust. In May’s world, robustness in GAAP enters through adaptation to the environment as it evolves. What does a theory of robustness look like that relies on flexibility and adaptation? Evolutionary game theory seems an obvious place to start.

A puzzle in capital budgeting is that firms continue to employ methods such as payback and internal rate of return that are suboptimal according to (neo)classical finance theory, which prescribes the NPV method. These other methods are often used as a secondary method to supplement NPV. Using multiple evaluation criteria seems a robust approach. As Ross (1995) writes, “[b]ecause the true NPV is unknown, the astute financial manager seeks clues to assess whether the estimated NPV is reliable. For this reason, firms would typically use multiple criteria for evaluating a project . . . [i]f different indicators seem to agree it’s ‘all systems go.’”

Another potential application is bonus caps. Are bonus caps common in practice because they are robust, for example, to information asymmetries about the tails of probability distributions? A recent

focus of incentive practice is bonus banks and other features of incentive arrangements that allow for bonus “clawbacks.” These seem intended as attempts at more robust compensation.

5.3 Implicit Contracts

What would a stewardship model of accounting look like that relies on implicit contracting (Glover, 2012)? If accounting and explicit contracting are used to set the stage for implicit contracting (Arya et al., 1997a), which financial accounting choices and limits on the set of allowable choices best set the stage? Is there an important qualitative difference between a stewardship model with a single steward and a model with a stewardship team that can mutually monitor and implicitly contract with each other?

One problem in financial reporting regulation is that preparers are aided by financial engineers and other advisors in their attempts to circumvent accounting standards. As soon as a standard is promulgated, an army of financial engineers devises a way around the new standard. The standard setters try to revise the rule or provide implementation guidance in response to the financial engineering, but those revisions and implementation guidance spawn new financial engineering. As the regulators struggle to keep pace, they are usually left in the dust. One way to discourage financial engineering designed to meet the letter of the rules but circumvent their spirit is a broad sense of mutual accountability — using team incentives (in this case punishments) to encourage mutual monitoring and discourage unwanted collaborations. If showing up on the customer list of a financial engineer found by a regulator to be selling an abusive product means the preparer will also be targeted for investigation, the preparer may run the other way when approached by a financial engineer about such a product (or even one who has a reputation for selling such products).

Zimmerman (2000, p. 330) suggests that cost allocation can have a role in providing team incentives. Noninsulating cost allocations, under which one manager benefits from the other’s good performance and larger allocation of a common fixed cost, can foster incentives for helping and mutual monitoring. In general, which managerial accounting

practices (alone or in combination) foster mutual monitoring, providing both an incentive for mutual monitoring and a means of enforcement?

In corporate governance, interlocking boards are criticized for facilitating collusion among managers and board members against shareholders. Can interlocking boards also facilitate mutual monitoring and implicit contracting for the benefit of shareholders?

Bonus pools are an example of a practice that seems to have the potential to provide a better understanding of the way in which incentives for cooperation among managers are provided for some actions, while incentives for competition are provided for other actions. Is group incentives (e.g., Arya et al., 1997a) a good model for helping and other mutually observable actions and individual incentives (e.g., Holmstrom, 1979) a good model for actions that managers take that are not mutually observable?

Most incentive theory focuses on single-period relationships and explicit contracting. Although research on implicit contracting has been on the rise, we mostly have results on fairly specific settings. Ichiishi (1993) present a more general model. He views the firm as best described by a hybrid solution concept. A manager agrees to play cooperatively (e.g., according to a core-theoretic behavioral principle) with some subset of managers (e.g., those within the same firm) and noncooperatively with others. If we are to use Ichiishi's view of the firm to understand observed institutions, it seems important to model the institutions that support cooperation rather than assuming cooperation directly. Implicit incentives that arise in repeated relationships and time spent together, which facilitates mutual monitoring and trust, seem likely to be key ingredients.

Many of the behavioral explanations for experimental results that do not conform to the predictions of one-shot game theory can be seen as having their foundation in repeated play (and evolutionary selection that favors players who are good at playing repeated games). Even in the simple two-period model of Arya et al. (1997a), two types of punishments emerge. Under strategic substitutes, the subgame perfect punishment can be interpreted as distrust. Under strategic complements, the subgame perfect punishment might instead be interpreted as revenge. Both distrust and revenge are usually viewed as undesirable.

Confucius wrote: before you embark on a journey of revenge, dig two graves. Yet, distrust and even revenge can also be viewed as useful threats that support the greater good of the team, firm, or even society as a whole, as in Adam Smith's *Theory of Moral Sentiments*. Before Adam Smith, the poet Edward Young wrote in 1721 "[w]hat is revenge, but courage to call in our honor's debts, and wisdom to convert others' self-love into our own protection?" Presumably, distrust is easier to foster than revenge. Which institutions are designed to foster distrust and/or revenge?

A

Appendix

Proof of Proposition 2.1. $\Pr(L|L)t_{LL}^* + \Pr(H|L)t_{LH}^* = c_L X$ and $\Pr(L|H)t_{HL}^* + \Pr(H|H)t_{HH}^* = c_H X$, which ensures the first-best payoffs are obtained. The contract satisfies dominant strategy incentives because each manager's report of his cost is used only in determining the other manager's transfers. \square

Proof of Proposition 3.1. Let λ_{IC} be the Lagrange multiplier on the (IC) constraint and λ_{km} be the multipliers on the bankruptcy (B) constraints. The solution is as specified in Proposition 3.1 with multipliers $\lambda_{IC} = \frac{p_H}{p_H - p_L}$, $\lambda_{LL} = (1 - p_H)(1 - \sigma)$, $\lambda_{LH} = p_H(1 - \sigma)$, $\lambda_{HL} = 0$, and $\lambda_{HH} = \sigma$. All of the multipliers are nonnegative. \square

Proof of Proposition 3.3. The added constraint is: $E(t|H, L) - H \geq E(t|L, L) - L$. Consider manager A. Suppose t^A does not depend on y^B . Call these payments t^{AU} for unconditional. Replace t^{AU} with payments conditioned on y^B by setting $t^A = t^{AU}/(1 - q_L^B)$. These new payments are positive if and only if $y^B = L$. This revision to the payments leaves $E(t|H, L)$ and $E(t|L, L)$ unchanged but reduces the equilibrium expected payment $E(t|H, H)$, since $(1 - q_H^B)/(1 - q_L^B) < 1$. So, the unconditional payments were not optimal. The only other

alternative is to condition manager A 's payments so that t^A is positive if and only if $y^B = H$. This would increase $E(t|H,H)$, since $q_H^B/q_L^B > 1$. \square

Proof of Proposition 3.4. The minimum IPE bonus is the solution to the revised incentive compatibility constraint as an equality, taking headquarters' intervention strategy as given: $p_H t_S^{IPE} + (1 - p_H)(1 - p_H)t_S^{IPE} - H \geq p_L t_S^{IPE} + (1 - p_L)(1 - p_H)t_S^{IPE} - L$, or $t_S^{IPE} = \frac{H}{p_H(p_H - p_L)}$.

Headquarters' bailout creates an endogenous correlation in the managers' environments and a multiple equilibrium problem. In particular, if manager j chooses H as the above individual performance evaluation contract assumes, then manager i 's choice of H instead of L increases the probability he will receive the bonus by $p_H + (1 - p_H)(1 - p_H) - [p_L + (1 - p_L)(1 - p_H)] = p_H(p_H - p_L)$. If instead, manager j chooses L , then manager i 's choice of H instead of L increases the probability he will receive the bonus by $p_L(p_H - p_L)$. Since $p_H(p_H - p_L) > p_L(p_H - p_L)$, (L, L) is also a Nash equilibrium. Headquarters' bailout creates a strategic complementarity in the managers' payoffs under IPE. Under the (H, H) equilibrium, each manager's utility is $\frac{[1 - p_H(1 - p_H)]H}{p_H(p_H - p_L)} - H$. Under the (L, L) equilibrium, each manager's utility is $\frac{[1 - p_L(1 - p_L)]H}{p_H(p_H - p_L)} \cdot \frac{[1 - p_H(1 - p_H)]H}{p_H(p_H - p_L)} - H < \frac{[1 - p_L(1 - p_L)]H}{p_H(p_H - p_L)}$ is equivalent to $\frac{p_H}{p_L} > \frac{(1 - p_L)}{(1 + p_H)}$, which is always true. So, the (L, L) equilibrium Pareto-dominates the (H, H) equilibrium in the managers' subgame.

Because of the managers' risk neutrality, RPE is also an optimal solution without headquarters' bailout decision. RPE is not affected by the bailout option so continues to be optimal. \square

Proof of Proposition 3.5. It suffices to add two new (IC) constraints, corresponding to ε and $-\varepsilon$. Let λ_{IC} , $\lambda_{IC\varepsilon}$, and $\lambda_{IC-\varepsilon}$ be the Lagrange multiplier on the three (IC) constraints and λ_{km} be the multipliers on the bankruptcy (B) constraints; $k, m = F, S$. Solving for the optimal contract assuming $\lambda_{IC\varepsilon}$ is the only positive (IC) multiplier, the optimal solution that emerges is RPE. The other (IC) constraints are also satisfied under this solution, so it is always feasible. Under the RPE solution, the key multiplier is $\lambda_{SS} = \frac{\sigma(1 - p_H) - \varepsilon(p_H + \sigma(1 - p_H))}{1 - p_H - \varepsilon}$, which

becomes negative under the condition given in the statement of proposition. Once ε exceeds that limit, IPE emerges as optimal. \square

Proof of Proposition 4.2. The optimal contract can take one of two possible approaches. The first is to rely on individual incentives. From Proposition 3.1, the optimal such contract is RPE.

The second possibility is to rely on the managers to mutually monitor each other in the first period and to provide them with the means of punishing each other in the second period. The optimal such first-period contract is to provide group incentives and reward only a success from both managers t_{SS}^G such that: $[\sigma + (1 - \sigma)p_H p_H]t_{SS}^G - H \geq [\sigma + (1 - \sigma)p_L p_L]t_{SS}^G - L$. Solving this equation as an equality implies $t_{SS}^G = \frac{H}{(1-\sigma)(p_H p_H - p_L p_L)}$. The benefit to each manager of free-riding under t_{SS}^G is $H - (p_H p_H - p_H p_L) \frac{H}{(1-\sigma)(p_H p_H - p_L p_L)}$. The least costly way of providing a punishment in the second period is to use a mix of IPE and JPE, where the JPE component provides the punishment. Define $t_S^{IPE} = \frac{H}{(1-\sigma)(p_H - p_L)}$ and $t_{SS}^{IPE} = \frac{H}{(1-\sigma)(p_H p_H - p_L p_H)}$. The weight on JPE that exactly offsets the first-period free-riding benefit is $w = \frac{p_H}{p_H + p_L}$. Comparing the objective functions of the two approaches yields the condition presented in the proposition. \square

Proof of Proposition 4.3. Start with the strategic complements case: $p_H - p > p - p_L$. The optimal contract provides the managers with incentives to mutually monitor each other, as in the two-period version of the model. JPE has to serve a second role in the infinitely repeated relationship: the gain from free-riding this period must be less than the infinite punishment of playing (L, L) instead of (H, H) in all future periods. So, t_S must satisfy: $pt_S - (p_H t_S - H) \leq \frac{p_H t_S - H - p_L t_S}{r}$. The smallest such t_S is $t_S^{\text{Cooperative}} = \frac{H(1+r)}{p_H - p_L + r(p_H - p)}$, which is (i). As the discount rate goes to 0, $t_S^{\text{Cooperative}}$ goes to $H/(p_H - p_L)$, which was our earlier first-period contract, or (iii).

In the case of strategic substitutes ($p_H - p < p - p_L$), there is a form of mutual monitoring that can emerge as equilibrium behavior that headquarters finds undesirable. Namely, the managers may find alternating between (H, L) and (L, H) more appealing than playing (H, H) in each period. They may use the stage game equilibrium of

(L, L) as a threat to enforce the alternating equilibrium rather than the (H, H) equilibrium. The manager who chooses H in the current period when the other manager chooses L has the lowest discounted expected utility: $\frac{(1+r)pt_S}{r} - \frac{H(1+r)^2}{r(2+r)}$. The (H, H) equilibrium must provide a higher discounted expected utility: $\frac{(1+r)(p_H t_S - H)}{r}$. The smallest such bonus is: $t_S = \frac{H}{(2+r)(p_H - p)}$, which proves (ii). As r goes to 0, the optimal JPE contract goes to $H/2(p_H - p)$, which is greater than in the complements case, or (iv). \square

Acknowledgments

I thank Tim Baldenius (the referee), Hui Chen, Pingyang Gao, Christian Hofmann, Xu Jiang, Bjorn Jorgensen, Wei Li, Stefan Reichelstein (the Editor), Kristy Towry, Hao Xue, and workshop participants at the University of Colorado at Boulder, Emory University, and the 2012 EIASM Workshop on Accounting and Economics in Segovia for their helpful comments. This monograph is based largely on my previous research. I thank my co-authors of the related papers: Tim Baldenius, Joel Demski, John Fellingham, Jack Hughes, Yuji Ijiri, Carolyn Levine, Pierre Jinghong Liang, Uday Rajan, K. Sivaramakrishnan, Shyam Sunder, Hao Xue, Richard Young, and particularly Anil Arya. I also thank David Schmeidler for his excellent course on mechanism design and subsequent discussions on “simpler” mechanisms and “approximate” (not “virtual”) implementation and Matthew Jackson for his feedback on that work.

References

- Abreu, D. and H. Maskava (1992), ‘Virtual implementation in iteratively undominated strategies: Complete information’. *Econometrica* **60**, 993–1008.
- Acharya, V. and T. Yorulmazer (2007), ‘Too many to fail — an analysis of time-inconsistency in bank closure policies’. *Journal of Financial Intermediation* **16**, 1–31.
- Adams, R. and D. Ferreira (2007), ‘A theory of friendly boards’. *The Journal of Finance* **62**, 217–250.
- Aggarwal, R. and A. Samwick (1999), ‘Executive compensation, strategic competition, and relative performance evaluation: Theory and evidence’. *The Journal of Finance* **54**, 1999–2043.
- American Institute of Certified Public Accountants (1934), *Audits of Corporate Accounts*. New York, NY.
- Antle, R. (1984), ‘Auditor independence’. *Journal of Accounting Research* **22**, 1–20.
- Antle, R. and J. Demski (1988), ‘The controllability principle in responsibility accounting’. *The Accounting Review* **63**, 700–718.
- Antle, R. and G. Eppen (1985), ‘Capital rationing and organizational slack in capital budgeting’. *Management Science* **31**, 163–174.

- Antle, R. and J. Fellingham (1990), 'Resource rationing and organizational slack in a two-period model'. *Journal of Accounting Research* **28**, 1–24.
- Antle, R. and J. Fellingham (1997), 'Models of capital investments with private information and incentives: A selective review'. *Journal of Business Finance and Accounting* **24**, 887–908.
- Antle, R. and A. Smith (1986), 'An empirical investigation of the relative performance evaluation of corporate executives'. *Journal of Accounting Research* **24**, 1–39.
- Arya, A., J. Demski, J. Glover, and P. Liang (2009), 'Quasi-robust multi-agent contracts'. *Management Science* **55**(5), 752–762.
- Arya, A., J. Fellingham, and J. Glover (1997a), 'Teams, repeated tasks, and implicit incentives'. *Journal of Accounting and Economics* **23**, 7–30.
- Arya, A., J. Fellingham, J. Glover, and K. Sivaramakrishnan (2000a), 'Capital budgeting, the hold-up problem, and information system design'. *Management Science* **46**, 205–216.
- Arya, A. and J. Glover (1995), 'A simple forecasting mechanism for moral hazard settings'. *Journal of Economic Theory* **66**, 507–521.
- Arya, A. and J. Glover (1996), 'The role of budgeting in eliminating tacit collusion'. *Review of Accounting Studies* **1**, 191–205.
- Arya, A. and J. Glover (1997), 'Maintaining auditor independence'. Working Paper, Carnegie Mellon University.
- Arya, A. and J. Glover (2001), 'Option value to waiting created by a control problem'. *Journal of Accounting Research* **39**, 405–415.
- Arya, A. and J. Glover (2006), 'Bailouts and unwanted coordination'. *Journal of Accounting, Auditing, and Finance* **21**, 109–117.
- Arya, A., J. Glover, and J. Hughes (1997b), 'Implementing coordinated team play'. *Journal of Economic Theory* **74**, 218–232.
- Arya, A., J. Glover, and U. Rajan (2000b), 'Implementation in principal-agent models of adverse selection'. *Journal of Economic Theory* **93**, 87–109.
- Arya, A., J. Glover, and S. Sivaramakrishnan (1997c), 'The interaction between decision and control problems'. *The Accounting Review* **72**, 561–574.

- Arya, A., J. Glover, and S. Sunder (1998), 'Earnings management and the Revelation Principle'. *Review of Accounting Studies* **3**, 7–34.
- Arya, A., J. Glover, and R. Young (1995), 'Virtual implementation in separable Bayesian environments using simple mechanisms'. *Games and Economic Behavior* **9**, 127–138.
- Arya, A., J. Glover, and R. Young (1996), 'Capital budgeting in a multidivisional firm'. *Journal of Accounting Auditing and Finance* **11**, 519–534.
- Baiman, S. (1975), 'The evaluation and choice of internal information systems within a multiperson world'. *Journal of Accounting Research* **13**, 1–15.
- Baiman, S., P. Fischer, M. Rajan, and R. Saouma (2007), 'Resource allocation auctions within firms'. *Journal of Accounting Research* **45**, 915–946.
- Baiman, S. and M. Rajan (1995), 'The informational advantages of discretionary bonus schemes'. *The Accounting Review* **70**(October), 557–579.
- Baker, G., R. Gibbons, and K. Murphy (1994), 'Subjective performance measures in optimal incentive contracts'. *Quarterly Journal of Economics* **109**(November), 1125–1156.
- Baldenius, T., S. Dutta, and S. Reichelstein (2007), 'Cost allocation for capital budgeting decisions'. *The Accounting Review* **82**, 837–867.
- Baldenius, T. and J. Glover (2012), 'Relational contracts with and between agents'. Working paper, Carnegie Mellon University and Columbia University.
- Bebchuk, L. and J. Fried (2006), *Pay Without Performance*. Boston, MA: Harvard University Press.
- Benoit, J. and V. Krishna (1985), 'Finitely repeated games'. *Econometrica* **53**, 905–922.
- Bergemann, D. and S. Morris (2005), 'Robust mechanism design'. *Econometrica* **73**, 1521–1534.
- Bergemann, D. and S. Morris (2008), 'Ex post implementation'. *Games and Economic Behavior* **63**, 527–566.
- Bernheim, B. D. and M. D. Whinston (1998), 'Incomplete contracts and strategic ambiguity'. *American Economic Review* **88**, 902–932.

- Bull, C. (1987), 'The existence of self-enforcing implicit contracts'. *The Quarterly Journal of Economics* **102**(February), 147–159.
- Bushman, R., Z. Dai, and W. Zhang (2012), 'Management team incentive alignment and firm value'. Working Paper, University of North Carolina at Chapel Hill.
- Che, Y. and S. Yoo (2001), 'Optimal incentives for teams'. *American Economic Review* pp. 525–541.
- Christensen, J. A. and J. S. Demski (2003), *Accounting Theory: An Information Content Perspective*. New York, NY: McGraw-Hill.
- Chung, K. and J. Ely (2007), 'Foundations of dominant strategy mechanisms'. *Review of Economic Studies* **74**, 447–476.
- Cremer, J. and R. P. McLean (1988), 'Full extraction of the surplus in Bayesian and dominant strategy auctions'. *Econometrica* **56**, 1247–1257.
- Demski, J. (2010), *Managerial Uses of Accounting Information*. New York, NY: Springer Verlag.
- Demski, J. and D. Sappington (1984), 'Optimal incentive contracts with multiple agents'. *Journal of Economic Theory* **33**(June), 152–171.
- Dutta, S. and X. Fan (2012), 'Incentives for innovation and centralized versus delegated capital budgeting'. *Journal of Accounting and Economics* **53**, 592–611.
- Dutta, S. and S. Reichelstein (2002), 'Controlling investment decisions: Depreciation and capital charges'. *Review of Accounting Studies* **7**, 253–281.
- Ederhof, M., M. Rajan, and S. Reichelstein (2010), 'Discretion in managerial bonus pools'. *Foundations and Trends in Accounting* **5**(4), 243–316.
- Gibbard, A. (1973), 'Manipulation of voting schemes: A general result'. *Econometrica* **41**, 587–602.
- Gibbons, R. and K. Murphy (1992), 'Optimal incentive contracts in the presence of career concerns: Theory and evidence'. *Journal of Political Economy* **100**, 468–505.
- Gibbs, M., K. Merchant, W. Van der Stede, and M. Vargus (2004), 'Determinants and effects of subjectivity in incentives'. *The Accounting Review* **79**(April), 409–436.

- Glover, J. (1994), 'A simpler mechanism that stops agents from cheating'. *Journal of Economic Theory* **62**, 221–229.
- Glover, J. (2004), 'A discussion of 'a model of auditing under bright-line accounting standards''. *Journal of Accounting, Auditing and Finance* **19**, 561–564.
- Glover, J. (2012), 'Disclosure and incentives'. *Accounting Horizons* **26**, 371–380.
- Glover, J., Y. Ijiri, C. Levine, and P. Liang (2005), 'Verifiability and manipulability vs. information asymmetries about verifiability and manipulability'. Working Paper, Carnegie Mellon University.
- Glover, J. and H. Xue (2012), 'A multi-period foundation for bonus pools'. Working Paper, Carnegie Mellon University.
- Hearn, L. (1924), *Interpretations of Literature*. New York, NY: Dodd, Mead and Company.
- Holmstrom, B. (1979), 'Moral hazard and observability'. *Bell Journal of Economics* **10**(Spring), 74–91.
- Ichiishi, T. (1993), *The Cooperative Nature of the Firm*. Cambridge: Cambridge University Press.
- Ijiri, Y. (1975), *Theory of Accounting Measurement*. Sarasota, FL: American Accounting Association.
- Itoh, H. (1993), 'Coalitions, incentives, and risk sharing'. *Journal of Economic Theory* **60**, 410–427.
- Jackson, M. (1992), 'Implementation in undominated strategies: A look at bounded mechanisms'. *Review of Economic Studies* **59**, 757–775.
- Jackson, M. and T. Palfrey (2001), 'Voluntary implementation'. *Journal of Economic Theory* **98**, 1–25.
- Janakiraman, S., R. Lambert, and D. Larcker (1992), 'An empirical investigation of the relative performance evaluation hypothesis'. *Journal of Accounting Research* **30**, 53–69.
- Jensen, M. and K. Murphy (1990), 'Performance pay and top-management incentives'. *Journal of Political Economy* **98**, 225–264.
- Laffont, J. and D. Martimort (2001), *The Theory of Incentives: The Principal-Agent Model*. Princeton, NJ: Princeton University Press.
- Lerrick, A. and A. Meltzer (2001), 'Beyond IMF bailouts: Default without disruption'. *Quarterly International Economics Report* **May 2001**.

- Levin, J. (2003), 'Relational incentive contracts'. *American Economic Review* **93**(June), 835–847.
- Li, C. (2012), 'Mutual monitoring within top management teams: A structural model investigation'. Working Paper, Carnegie Mellon University.
- Ma, C. (1988), 'Unique implementation of incentive contracts with many agents'. *Review of Economic Studies* **55**, 555–572.
- Ma, C., J. Moore, and S. Turnbull (1988), 'Stopping agents from "cheating"'. *Journal of Economic Theory* **46**, 355–372.
- MacLeod, W. (2003), 'Optimal contracting with subjective evaluation'. *American Economic Review* **93**(March), 216–240.
- Mautz, R. and H. Sharaf (1961), *The Philosophy of Auditing*. Sarasota, FL: American Accounting Association.
- May, G. O. (1958), 'Generally accepted principles of accounting'. *Journal of Accountancy* **105**(1), 23–27.
- Milgrom, P. and J. Roberts (1992), *Economics, Organizations and Management*. Englewood Cliffs, NJ: Prentice Hall.
- Mookherjee, D. (1984), 'Optimal incentive schemes with many agents'. *Review of Economic Studies* **51**, 433–446.
- Mookherjee, D. and S. Reichelstein (1990), 'Implementation via augmented revelation mechanisms'. *Review of Economic Studies* **57**, 453–475.
- Mookherjee, D. and S. Reichelstein (1992), 'Dominant strategy implementation of Bayesian incentive compatible rules'. *Journal of Economic Theory* **56**, 378–399.
- Murphy, K. and P. Oyer (2003), 'Discretion in executive incentive contracts'. Working Paper, University of Southern California and Stanford University.
- Myerson, R. (1979), 'Incentive compatibility and the bargaining problem'. *Econometrica* **47**, 61–74.
- Rajan, M. (1992), 'Cost allocation in multiagent settings'. *The Accounting Review* **67**, 527–545.
- Rajan, M. and S. Reichelstein (2006), 'Subjective performance indicators and discretionary bonus pools'. *Journal of Accounting Research* **44**(June), 585–618.

- Rajan, M. and S. Reichelstein (2009), 'Objective versus subjective indicators of managerial performance'. *The Accounting Review* **84**(January), 209–237.
- Reichelstein, S. (1997), 'Investment decisions and managerial performance evaluation'. *Review of Accounting Studies* **2**, 157–180.
- Riordan, M. and D. Sappington (1988), 'Optimal contracts with public ex post information'. *Journal of Economic Theory* **45**, 189–199.
- Ross, S. (1995), 'Uses, abuses, and alternatives to the net-present-value rule'. *Financial Management* (Autumn), 96–102.
- Satterthwaite, M. (1975), 'Strategy-proofness and Arrow's conditions: Existence and correspondence theorems'. *Journal of Economic Theory* **10**, 187–217.
- Schmeidler, D. (1980), 'Walrasian analysis via strategic outcome functions'. *Econometrica* **48**, 1585–1593.
- Sunder, S. (1997), *Theory of Accounting and Control*. Cincinnati, OH: South-Western Publishing.
- Tirole, J. (1986), 'Hierarchies and bureaucracies: On the role of collusion in organizations'. *Journal of Law, Economics and Organization* **2**, 181–214.
- Tirole, J. (1992), 'Collusion and the theory of organizations'. In: J.-J. Laffont (ed.): *Advances in Economic Theory: Invited Papers for the Sixth World Congress of the Econometric Society 2*. Cambridge: Cambridge University Press.
- Towry, K. (2003), 'Control in a teamwork environment — the impact of social ties on the effect of mutual monitoring contracts'. *The Accounting Review* **78**, 1069–1095.
- Wilson, R. (1987), 'Game-theoretic analysis of trading processes'. In: T. Bewley (ed.): *Advances in Economic Theory (Fifth World Congress)*.
- Zimmerman, J. (2000), *Accounting for Decision Making and Control*. New York, NY: McGraw-Hill.