Econometrica, Vol. 74, No. 4 (July, 2006), 1063–1107

ROBUSTLY COLLUSION-PROOF IMPLEMENTATION

BY YEON-KOO CHE AND JINWOO KIM [1]

A contract with multiple agents may be susceptible to collusion. We show that agents' collusion imposes no cost in a large class of circumstances with risk neutral agents, including both uncorrelated and correlated types. In those circumstances, any payoff the principal can attain in the absence of collusion, including the second-best level, can be attained in the presence of collusion in a way robust to many aspects of collusion behavior. The collusion-proof implementation generalizes to a setting in which only a subset of agents may collude, provided that noncollusive agents' incentives can be protected via an ex post incentive compatible and ex post individually rational mechanism. Our collusion-proof implementation also sheds light on the extent to which hierarchical delegation of contracts can optimally respond to collusion.

KEYWORDS: Robustly collusion-proof implementation, pairwise identifiability, subgroup collusion, hierarchical delegation.

1. INTRODUCTION

THERE HAS BEEN A GROWING INTEREST in studying collusion among asymmetrically informed agents, in various settings ranging from internal organization, regulation, and auctions, to oligopolistic competition.[2] Although most of these studies explore how agents can effectively collude against exogenously given institutions, a few recent studies have begun to investigate an *optimal* organizational/contractual response to agents' collusion. In particular, Laffont and Martimort (1997, 2000) have developed a modeling framework that integrates collusion as part of the general mechanism design analysis.[3] An important insight gained from this framework is that agents' asymmetric information imposes transaction costs on their abilities to carry out collusive arrangements.

[1]The authors are grateful for comments and suggestions from a co-editor and three referees, and from Gorkem Celik, Jacques Crémer, Ian Gale, Lucía Quesada, Larry Samuelson, Bill Sandholm, Sergei Severinov, Guofu Tan, and seminar participants at USC, UCSD, Korea University, and Yonsei University. The authors acknowledge warm hospitality they received from Yonsei University and the University of Wisconsin, respectively, during their visits. The authors have benefited from financial support from the Shoemaker Fellowship and from Yonsei University.

[2]Tirole (1986), Baliga and Sjöström (1998), Celik (2004), Faure-Grimaud, Laffont, and Martimort (2003), Severinov (2003), and Mookherjee and Tsumagari (2004) study collusion in internal organization and the value of delegation. Graham and Marshall (1987), McAfee and McMillan (1992), Mailath and Zemsky (1991), Marshall and Marx (2004), Brusco and Lopomo (2002), Caillaud and Jehiel (1998), and Esö and Schummer (2004) study collusion in one-shot auctions of various formats, while Aoyagi (2003), Blume and Heidhues (2002), Skrzypacz and Hopenhayn (2004), and Abdulkadiroğlu and Chung (2003) study collusion in repeated auctions.

[3]Earlier literature concerned about coalition formation in Groves' mechanisms includes Green and Laffont (1979) and Crémer (1996). The former paper envisions a coalition of symmetrically informed agents, whereas the latter allows for their possible asymmetric information. Although the latter framework resembles that of Laffont and Martimort, and even considers subgroup collusion, it restricts attention to dominant strategy implementation (at both grand and coalitional mechanism design) and does not consider participation constraints.

Just how far these transaction costs can be exploited in contract design is still unknown. In procurement/public good settings, Laffont and Martimort (hereafter LM) have shown that the optimal outcome can be made collusion-proof at no cost to the principal if the agents' types are uncorrelated (LM, 1997), but if the types are correlated, preventing collusion entails strict cost to the principal (LM, 2000). The former result—i.e., collusion is preventable at no cost with uncorrelated types—is reproduced by Quesada (2004) with a different coalition formation process, and by Jeon and Menicucci (2005) in a nonlinear pricing model that allows collusive consumers to arbitrage on their purchases. These models have special structures, though. Laffont and Martimort and Quesada (2004) assume two agents with two possible types and Leontief production technologies/preferences, and Jeon and Menicucci (2005) assume $n \geq 2$ agents with two types or two agents with three types, along with several preference restrictions.

Intriguing as these results are, their reliance on special structures raises several questions. First, it is unclear whether the results are generalizable beyond the assumed environments. Second, even if the results are generalizable, the method of collusion-proof implementation is specific to the assumed setting, so it does not provide a general method that may work in other settings. Third, the specificity of the models and the lack of a general method also make it difficult to isolate the economic insight that explains under what circumstances collusion is preventable and why it is preventable in those circumstances.

The current paper advances on these fronts by developing a general method for collusion-proofing a mechanism. Using this method, we show that any payoff attainable by the principal in the absence of collusion, including the second-best level, can be attained in the presence of collusion in a large class of environments with risk neutral agents, for both uncorrelated and correlated types cases. Our collusion-proof implementation does not rest on any special assumptions about preferences/technologies or type structures. For example, the agents' types can be discrete or continuous (at least for the uncorrelated types case) or even multidimensional, and no special features on preferences or technology, such as single crossing, are needed for our results.

Furthermore, our collusion-proof implementation is robust to many aspects of collusion behavior, such as the identity of the agent organizing/initiating collusion, the manipulation technology employed by the coalition (e.g., whether the coalition can arbitrage on an initial allocation), the coalition's objective and the bargaining power of its members (e.g., whether the coalition caters to the interests of some agents more than others), and the exact makeup of the coalition (e.g., whether collusion involves all agents or only some agents). In fact, the principal need not even know how the collusion operates along many of these dimensions.

Our method of collusion-proof implementation utilizes the idea of "selling the firm to the coalition." Specifically, for any expected payoff level that the principal can attain in the absence of collusion, we construct a new mechanism that gives the principal an ex post constant payoff equal to the original expected payoff. This mechanism forces the (grand) coalition to become a residual claimant of the entire surplus, after paying off the principal an ex post constant surplus, when it manipulates the outcome. That such a mechanism is implementable in the adverse selection setting is not obvious and will be an important part of our analysis. Also not obvious is that such a mechanism, if implementable, is immune to collusion. In fact, being the residual claimant, the coalition would prefer the first-best allocation over the intended allocation in case the latter involves distortion, so it will try to manipulate so that the former allocation arises. Yet, such a manipulation never succeeds. The reason is that the coalition faces an asymmetric information problem just like the principal in the original noncollusive mechanism design. This informational asymmetry means that an appropriate amount of information rent must be given to the members of the coalition to implement a particular allocation. However, since the principal is paid off to realize a desired level of surplus irrespective of the induced allocation, implementing any other allocation by the coalition would violate budget balancing.[4] (This intuition will become more transparent in Section 5, with the aid of a figure.) In short, by making the agents residual claimants, our mechanism forces them to internalize precisely the same amount of informational cost that the principal faces in noncollusive mechanism design, and in this sense exploits the coalitional transaction cost fully.

This idea of collusion-proof implementation does not rely on the agents' types being uncorrelated, although making the agents residual claimants while preserving their incentives proves more challenging in a correlated type environment. If there are only two agents, our method of collusion-proofing indeed does not work, much consistent with LM's (2000) finding in their two agents model. With more than two agents, however, given a reasonable type structure, our collusion-proof implementation works quite generally, implying again that the principal can attain any noncollusive payoff in a robustly collusion-proof fashion even with correlation. An important corollary of this result is that the principal can typically implement the first-best allocation and extract the entire rents from the agents even in the presence of collusive agents.

We then extend our analysis to consider a mechanism that would prevent collusion by a subgroup of agents. Although the issue of preventing collusion by a subgroup has rarely been analyzed before, it is practically relevant because in many settings, only a subgroup of agents is often in a position to collude. Collusion-proofing in this environment poses a new challenge because the coalition may prey on noncollusive agents as much as on the principal. Protecting the interests of noncollusive agents thus becomes an important consideration for the principal. Our collusion-proof implementation idea generalizes in a remarkable way to this problem: If *at least two* collusive agents are

[4]The intuition is the same as the one showing that implementing the first-best allocation would run a budget deficit in Myerson–Satterthwaite (1983) bargaining. The difference is that this problem is endogenously/deliberately created by our design to prevent collusion from being feasible.

identified, then we can construct a mechanism that can handle *any* collusion that involves these two, including the grand collusion. This result strengthens the robustness in the way the collusion problem is thwarted since the principal need not know the exact size or makeup of the coalition. Although this result requires an additional condition that the outcome must be ex post implementable in the noncollusive setup (i.e., ex post incentive compatible and ex post individually rational), the condition is known to hold in a large class of uncorrelated types environments.

The collusion-proof implementation result also advances our understanding of the value of hierarchical delegation of contracts. Despite its practical significance, delegation of contracting authority has been difficult to justify, since it involves a loss of control for the principal (see Melumad, Mookherjee, and Reichelstein (1995), for instance). Whether collusion can change this view has been the subject of much recent research (see Laffont and Martimort (1998), Faure-Grimaud, Laffont, and Martimort (2003), Celik (2004), Mookherjee and Tsumagari (2004)). Since collusion creates control loss even with centralized contracts, delegation may be relatively more attractive and may even serve as an optimal response to collusion. This latter conjecture turns out not to be true, however. Our results imply that collusion imposes no real cost to centralized contracting, which suggests that delegation cannot be more justifiable in the presence of collusion than in its absence.

The rest of the paper is organized as follows. Section 2 illustrates the idea of the main results using a simple example. Section 3 describes the model, including the economic environments studied. Section 4 describes the noncollusive benchmark. Section 5 develops the notion of robust collusion-proofness. Section 6 constructs a robustly collusion-proof mechanism that implements any noncollusive payoff for the principal, in the uncorrelated type environment. Section 7 generalizes the analysis to the correlated type environment. Section 8 then studies collusion-proofing when only a subset of agents may collude. Section 9 establishes robustness of the result to an alternative modeling of coalition formation. Section 10 draws implications for hierarchical delegation of contracts. Section 11 is the conclusion.

2. AN ILLUSTRATIVE EXAMPLE

It is useful to begin with an example that illustrates our main idea. Suppose a buyer procures a good from one of two suppliers, agents 1 and 2. Agent $i = 1, 2$ can supply the good at a cost θ_i, which is drawn uniformly from $[0, 1]$, and the buyer values the good more than 2. If the agents cannot collude, it is optimal for the buyer to use a standard auction, such as a second-price auction. (No binding reserve price is employed since the seller's valuation of the object is sufficiently high.) Specifically, the agents bid supply prices, and the low bidder wins and performs the job at the payment that equals the high bid. Consequently, the buyer procures the good at the expected price of $2/3$, which is the best the buyer can do, as is well known from Myerson (1981).

Suppose now the agents can collude. It is easily seen that the second-price auction is susceptible to collusion. Prior to bidding, the firms can organize a knockout auction wherein the agents bid for the right to participate in the second-price auction uncontested; i.e., the loser bids 1 and the winner bids his cost.[5] Hence, with collusion, the buyer essentially pays the price of 1 to the winner of the knockout auction.

Now consider a different mechanism. The buyer holds an auction in which the agents bid for a payment b_i and again the low bidder wins. The mechanism differs in the payment arrangement: The buyer pays a fixed amount, 2/3, to the losing (high) bidder, say j, who then pays the winning bidder its bid b_i to perform the job. Intuitively, the losing bidder is a "prime contractor" who "outsources" the job to the winning bidder and in the process finances the difference, $b_i - 2/3$.

Absent collusion, the bidding game has a unique equilibrium in which the agents adopt a symmetric increasing bidding strategy $\frac{1}{2} + \frac{1}{3}\theta$ for each type $\theta \in [0, 1]$. Consequently, the job is allocated efficiently as in the optimal mechanism and the buyer procures the good at the fixed price of 2/3. Since the allocation is the same and the buyer pays the same on average as in the (noncollusive) second-price auction, the revenue equivalence theorem implies that the interim payoffs of both firms are the same as in that game. Hence, it is equilibrium for both agents to participate in the auction game. In sum, the proposed mechanism implements the optimal procurement policy, in the absence of collusion. More importantly, the new mechanism is not susceptible to collusion. In the bidding game, the agents become residual claimants of the social surplus after paying a fixed amount of 2/3 to the buyer. Since the allocation is efficient, they have no incentive to collude in that bidding game.

This example illustrates the main idea of preventing collusion, namely that of "selling the firm" to the agents. In what follows, this idea will be used to construct a general collusion-proof mechanism that works in a more complicated environment. The example also illustrates another feature of our collusion-proof mechanism, distinguished from the existing literature (e.g., LM (1997, 2000)). Unlike the traditional approach, our mechanism guarantees the buyer a desired level of ex post surplus, whether collusion actually occurs or not. Hence, in the example, the buyer could achieve the same outcome by delegating the procurement job to a "consortium" of agents (run by some uninformed third party) at a fixed price of 2/3; the consortium will then organize its own auction to allocate the job efficiently. Such delegation may provide a more practically relevant implementation of our mechanism.

[5]More precisely, they can organize a knockout auction in which the agents bid to pay their rivals for "uncontested bidding" in the official auction. This knockout auction game has a unique symmetric equilibrium in which an agent with cost θ bids $\frac{1}{3} - \frac{1}{3}\theta$. This equilibrium implements the direct revelation (strong) cartel mechanism studied by McAfee and McMillan (1992). A similar problem arises with the first-price auction.

3. PRIMITIVES

There are a principal and $n \geq 2$ agents, with $N := \{1, \dots, n\}$ representing the total set of agents. Each agent i has type θ_i drawn from some arbitrary measurable set Θ_i. A vector of realized types is denoted $\theta := (\theta_1, \dots, \theta_n) \in \times_{i=1}^{n} \Theta_i =: \Theta$. Until more specific cases are considered, we maintain a general assumption that θ is distributed according to some prior distribution $\mu^0 \in \Delta\Theta$. Hence, θ can be discrete or continuous (or a mixture of those), or multidimensional. The realized value of θ_i is private information observed only by agent i; all others, including the principal, only know its distribution along with other aspects of the game structure. We adopt the following notation: $\tilde{\theta}$, $\tilde{\theta}_i$, and $\tilde{\theta}_{-i}$ represent random variables; $\mathbb{E}[\cdot] := \int_{\Theta} [\cdot] \, d\mu^0(\tilde{\theta})$ and $\mathbb{E}_{\tilde{\theta}_{-i}}[\cdot] := \int_{\Theta_{-i}} [\cdot] \, d\mu^0(\theta_i, \tilde{\theta}_{-i})$ are expectation operators based on the prior distribution; and $\mathbf{E}_{\mu}[\cdot] := \int_{\Theta} [\cdot] \, d\mu(\tilde{\theta})$ represents an expectation operator based on an arbitrary probability distribution $\mu \in \Delta\Theta$.

An economic decision is described by $x \in \mathcal{X}$ for some arbitrary set $\mathcal{X}$. Given a profile of types $\theta \in \Theta$ and a decision $x \in \mathcal{X}$, agent $i \in N$ realizes a gross surplus of $a_i(x, \theta)$ and the principal obtains $w(x)$. We allow for a random decision, so we focus on a probability measure q on $\mathcal{X}$ and call it an *allocation*. Let $\mathcal{Q} = \Delta\mathcal{X}$ be the set of all allocations (i.e., all probability measures on $\mathcal{X}$). Then any allocation q (or randomization over x) yields a gross surplus of $s_i(q, \theta) := \int_{\mathcal{X}} a_i(x, \theta) \, dq(x)$ and $v(q) := \int_{\mathcal{X}} w(x) \, dq(x)$ to agent i and to the principal, respectively, given type profile $\theta \in \Theta$.

All players are risk neutral.[6] Hence, given types θ, if allocation $q \in \mathcal{Q}$ is chosen and the principal pays t_i to agent i in expected value, he receives expected payoff of

$$s_i(q, \theta) + t_i$$

and the principal receives expected payoff[7]

$$v(q) - \sum_{i \in N} t_i.$$

[6]As will be remarked, our results continue to hold even if the principal is risk averse.

[7]In several models including LM, the principal is a government agency that cares about the agents' welfare. In that case, the principal's payoff is described as

$$v(q) - \sum_{i \in N} t_i + \lambda \sum_{i \in N} [s_i(q, \theta) + t_i]$$

for some $\lambda \in (0, 1]$. This objective function is relevant for a public good problem or Baron and Myerson's (1982) regulation problem, where $\lambda > 0$ reflects the government's shadow value of firms' revenue. Our method works even in this case for the optimal noncollusive mechanism, but works according to LM's weak collusion-proofness criterion, which is weaker than the one that will be developed here. The precise notion and the result are discussed in Appendix A of our working paper version (Che and Kim (2004)).

If there is no contract, agent i with type θ_i collects a reservation utility of $\overline{U}_i(\theta_i)$.

Virtually all known adverse selection problems with "quasilinear preferences" satisfy the above preference and information structure. The following list details some well-known examples.

- *Internal Organization, Procurement, and Regulation*. An employer/regulator procures a set K of goods in varying quantities from a set N of workers/firms. The decision $x := (x_i^k)_{k\in K, i\in N}$ then represents vectors of goods supplied by the workers. This situation easily fits into our model, where $a_i(x_i, \theta_i)$ represents worker $i \in N$'s payoff (i.e., negative of cost) associated with supplying a vector of quantities $x_i = (x_i^k)_{k\in K} \geq \mathbf{0}$ given his realized type θ_i (which can be multidimensional) and $w(x)$ represents the seller's value of procuring x. An allocation then is a probability distribution over different production assignments.
- *Nonlinear Pricing.* A firm produces/markets a set of goods K in varying quantities to a set N of consumers. This is just the mirror image of the procurement problem, with a decision x that represents the bundles of goods consumed by the buyers, and with $a_i(x, \theta_i)$ that represents consumer i's utility from consumption and $w(x)$ that represents the negative of the firm's cost of producing x.
- *Auctions.* An auctioneer allocates a (finite) set of goods or procurement projects K to n bidders and possibly to herself. Let $\mathcal{X}$ be the set of all partitions, or "assignments," of K into the set of all players, including the auctioneer. Suppose that $a_i(x, \theta)$, $i \in N$, is bidder i's gross surplus and $w(x)$ is the auctioneer's gross surplus when partition $x \in \mathcal{X}$ is chosen and the bidders realize types θ. This model covers many situations of interest, ranging from a one-unit independent private value (IPV) (seller or buyer) auction as the simplest form, to interdependent valuations (seen by the possible dependence of a_i on θ_{-i}), bundling, and Jehiel–Moldovanu–Stacchetti (1999) type allocation externalities. In such a model, an allocation $q = (q_x)_{x\in\mathcal{X}}$ denotes a vector of probabilities of different partitions being chosen, and $s_i(q, \theta) = \sum_{x\in\mathcal{X}} q_x a_i(x, \theta)$, $i = 1, \dots, n$, and $v(q) = \sum_{x\in\mathcal{X}} q_x w(x)$.

To describe the sequence of events, it is useful to begin with a time line under a *noncollusive game*:

- At date -1, each agent learns his type, θ_i, which is drawn from Θ_i.
- At date 0, the principal proposes a mechanism (to be described fully).
- At date 1, each agent either accepts or rejects the mechanism.
- At date 2, the game form proposed in the mechanism is played if the agents all accepted the mechanism or else no mechanism is played and the agents collect their respective reservation utilities.

To study possible collusion among the agents, we follow LM (2000) by considering possible coalition formation *between* date 1 and date 2, initiated by a third party[8]:

- At date $1\frac{1}{4}$, a third party proposes a collusive arrangement.
- At date $1\frac{1}{2}$, each agent accepts or rejects the collusive mechanism.
- At date $1\frac{3}{4}$, the game form specified in the collusive mechanism is played, which binds the play of the coalition members at date 2, if all agents accepted the collusive mechanism at date $1\frac{1}{2}$. If at least one agent rejects the collusive mechanism, no collusion occurs, so the agents play the game at date 2 noncooperatively.

Note that the coalition is formed after the agents make participation decisions. The implication of this formulation will be discussed in the Conclusion. Further details of how collusion operates will be discussed in Section 5.

4. BENCHMARK: A NONCOLLUSIVE ENVIRONMENT

We first analyze the noncollusive game with no action between date 1 and date 2. Absent collusion by agents, the revelation principle guarantees that an equilibrium consequence of any contract that the principal offers can be studied by a direct revelation mechanism. In our setup, a direct mechanism (*mechanism* for short) consists of measurable functions $(q, t): \Theta \mapsto \mathcal{Q} \times \mathbb{R}^n$, which determine an allocation $q(\theta)$ and a vector of transfers $t(\theta) = (t_1(\theta), \dots, t_n(\theta))$ to the agents when they report $\theta \in \Theta$. The function $q(\cdot)$ is called an *allocation rule* and the function $t(\cdot)$ is called a *transfer rule*. Any such pair (q, t) also represents an outcome realized at each state θ and will be sometimes referred to as an *outcome* below.

Absent collusion, a mechanism $M = (q, t)$ is *feasible* if it is *individually rational*,

$$\text{(IR)} \qquad U_i^M(\theta_i) := \mathbb{E}_{\tilde{\theta}_{-i}}\big[s_i(q(\theta_i, \tilde{\theta}_{-i}), \theta_i, \tilde{\theta}_{-i}) + t_i(\theta_i, \tilde{\theta}_{-i}) | \theta_i\big] \ge \overline{U}_i(\theta_i) \quad \forall i, \theta_i,$$

and *incentive compatible*,

$$\text{(IC)} \qquad \begin{aligned} U_i^M(\theta_i) \ge\ & \mathbb{E}_{\tilde{\theta}_{-i}}\big[s_i(q(\theta_i', \tilde{\theta}_{-i}), \theta_i, \tilde{\theta}_{-i}) + t_i(\theta_i', \tilde{\theta}_{-i}) | \theta_i\big] \\ =:\ & u_i^M(\theta_i', \theta_i) \quad \forall i, \theta_i, \theta_i', \end{aligned}$$

where $\overline{U}_i(\theta_i)$ is the reservation utility level of agent i with type θ_i. Notice that both incentive compatibility and individual rationality are required at the interim level. Let $\mathcal{M}$ be the set of all feasible mechanisms, i.e., the set of all allocation and transfer rules $M := (q, t)$ that satisfy (IC) and (IR). We assume that the set $\mathcal{M}$ is nonempty.[9]

[8] In Section 9, we will study a variation in which collusion is initiated by one of the agents.

[9] It is reasonable in most situations that the principal has an option to offer a null contract, in which case this assumption holds trivially.

A mechanism $M = (q, t) \in \mathcal{M}$ *implements* an (expected) payoff of $V \in \mathbb{R}$ for the principal if

$$V = \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i \in N} t_i(\tilde{\theta})\right],$$

in which case we say a payoff V is *implementable*. Let $\mathcal{V}$ denote the set of all implementable payoffs for the principal. Of special interest is the highest implementable payoff $V^* = \sup \mathcal{V}$. This payoff, henceforth referred to as *noncollusive optimal* or *second-best* payoff, is implementable, namely, $V^* \in \mathcal{V}$ under very weak conditions (see Balder (1996), for example). Subsequently, we will be interested in the collusion-proof implementability of any arbitrary $V \in \mathcal{V}$, but particularly the second-best payoff V^*.

For any payoff $V \in \mathcal{V}$, there may be more than one mechanism that implements it. For the most part, how we select a mechanism in such a case does not matter. On a couple of occasions (Propositions 2 and 3), however, we select a mechanism $M = (q, t)$ that *efficiently implements* $V \in \mathcal{V}$ in the sense that M yields the highest total surplus among all feasible mechanisms that implement V:

$$\mathbb{E}\left[v(q(\tilde{\theta})) + \sum_{i \in N} s_i(q(\tilde{\theta}), \tilde{\theta})\right] \geq \mathbb{E}\left[v(q'(\tilde{\theta})) + \sum_{i \in N} s_i(q'(\tilde{\theta}), \tilde{\theta})\right]$$

for any mechanism $M' = (q', t')$ that implements V. The existence of such a mechanism for any given $V \in \mathcal{V}$ involves a restriction, but it is a very weak one.[10] In particular, the second-best allocation—the allocation rule that implements V^*—is often unique, in which case any optimal mechanism will implement V^* efficiently.

5. MODEL OF COLLUSION AND COLLUSION-PROOFNESS

The Laffont–Martimort model of collusion postulates that the agents can commit, via an uninformed benevolent representative, to a mechanism that manipulates their reports to the principal. Below, we expand this modeling framework to accommodate a much broader range of collusion possibilities. We then develop a notion of collusion-proofness that requires a mechanism to be robust against all such collusion possibilities.

5.1. *Modeling Collusive Behavior*

We study a collusive arrangement that allows the agents (i) to collectively manipulate their reports to the principal, (ii) to reallocate q assigned by the

[10]For instance, if the set of allocation rules associated with mechanisms that implement V is compact, then there exists a mechanism that efficiently implements V, since the principal and agents' payoff functions are linear in q.

grand contract, and (iii) to exchange transfers among the agents in a budget-balanced way. Following LM, we assume that such an arrangement is enforced by a side contract proposed by a benevolent representative. By the revelation principle, a side contract is described without loss of generality by a pair of functions $(\mu, y) : \Theta \mapsto \Delta\Theta \times \mathbb{R}^n$ that map from the agents' types into (possibly random) reports they will submit to the principal and side transfers they will exchange with one another. Specifically, a side contract (μ, y) asks the agents to report their types and, for any profile θ of reported types, it instructs them to randomize their reports over Θ according to a probability measure $\mu(\theta)$ and to exchange side transfers $y(\theta) = (y_1(\theta), \dots, y_n(\theta))$ among them. We require a side contract to be *budget balanced*: $\sum_{i\in N} y_i(\cdot) = 0$.[11]

For our purpose, it is more convenient to work directly with the outcome that is implemented *as a result of* enforcing a balanced-budget side contract. Given any grand mechanism M, we say a mechanism $\tilde{M} := (\tilde{q}(\cdot), \tilde{t}(\cdot))$ is a *reallocational manipulation of* M if there exists a balanced-budget side contract $(\mu, y) : \Theta \mapsto \Delta\Theta \times \mathbb{R}^n$ such that, for each $\theta \in \Theta$,

$$(1) \qquad \tilde{t}(\theta) = \mathbf{E}_{\mu(\theta)}[t(\tilde{\theta})] + y(\theta) \quad \text{and} \quad v(\tilde{q}(\theta)) = \mathbf{E}_{\mu(\theta)}\big[v(q(\tilde{\theta}))\big],$$

and we let $\mathcal{RM}_M$ denote the set of all reallocational manipulations of M. In words, a reallocational manipulation of M is any outcome $\tilde{M} = (\tilde{q}, \tilde{t})$ that the coalition can induce from grand contract M by manipulating the reports from θ via randomization $\mu(\theta)$ and reallocating the resulting assignment in any way that gives rise to the same gross surplus to the principal (the second equation of (1)), and by redistributing transfers to the agents in a budget-balanced way (the first equation of (1)).

It is worth noting that the second equation of (1) encompasses all standard scenarios of reallocation. In an auction, for instance, a bidding ring may be able to reallocate the goods among themselves after they are initially auctioned off by the seller. This power to reallocate matters only when the good is sold to one of the members, however. The equation captures (a weaker form of) this restriction. To be more concrete, consider a single-unit auction with n bidders and a seller with a reservation value of $v_0 \geq 0$. Suppose the seller wishes to implement an allocation rule $q(\cdot) = (q_1(\cdot), \dots, q_n(\cdot))$, where $q_i(\cdot)$ is the probability of the object being allocated to agent i (as a function of θ). If the bidders can reallocate the object once it is assigned, they can induce any $\tilde{q}(\theta) = (\tilde{q}_1(\theta), \dots, \tilde{q}_n(\theta))$ as long as $\sum_{i\in N} \tilde{q}_i(\theta) = \mathbf{E}_{\mu(\theta)}[\sum_{i\in N} q_i(\tilde{\theta})]$ for some $\mu(\theta) \in \Delta\Theta$; i.e., the probability of at least one of them getting the

[11]In fact, all results except Proposition 1 hold with a weaker ex ante version of budget balancedness, i.e., $\mathbb{E}[\sum_{i\in N} y_i(\theta)] = 0$. This means that all of our collusion-proof implementation method works, even when the coalition is allowed to obtain financing from outside the coalition. Likewise, our collusion-proof implementation of optimal mechanisms works even when the coalition is allowed to burn money, i.e., with a weaker requirement $\sum_{i\in N} y_i(\cdot) \leq 0$.

good matches that under some (possibly randomized) reports. However, this condition implies

$$v(\tilde{q}(\theta)) = v_0 \cdot \left(1 - \sum_{i \in N} \tilde{q}_i(\theta)\right) = v_0 \cdot \left(1 - \mathbf{E}_{\mu(\theta)}\left[\sum_{i \in N} q_i(\tilde{\theta})\right]\right)$$

$$= \mathbf{E}_{\mu(\theta)}\left[v(q(\tilde{\theta}))\right],$$

which is precisely what we require under reallocational manipulation. In another example, as Jeon and Menicucci (2005) or Mookherjee and Tsumagari (2004) envisioned, consumers who face nonlinear pricing or suppliers who face nonlinear contracts may be able to reallocate their initial allocation/assignment to increase their joint surplus. In this case, our equation corresponds to the restriction that the reallocation cannot affect the total amount of the goods/outputs being (re)allocated to all consumers/suppliers.

For feasible collusive behavior, we focus on a reallocational manipulation that satisfies (IC) and (IR), and we let

$$\mathcal{M}_M := \mathcal{RM}_M \cap \mathcal{M}$$

be the set of *feasible* (reallocational) manipulations. Conditions (IC) and (IR) are sensible properties to assume for coalitional manipulation. First of all, (IC) is necessary as long as the coalition faces an adverse selection problem, regardless of how the coalition is formed. For instance, if the coalition is proposed by an (uninformed or informed) agent, the proposal must be incentive compatible for all agents, including the proposer (see Quesada (2004) and Mookherjee and Tsumagari (2004)). Likewise, (IR) is necessary for a collusion proposal to be acceptable to the agents in many circumstances. Whether a particular collusion proposal is acceptable depends on the belief formed when the proposal is (unexpectedly) rejected. A standard treatment for this is to assume "passivity of beliefs," i.e., no new inferences about the agents' types are made in such an event. Given passive beliefs, a manipulation $\tilde{M}$ would be acceptable if

$$(\mathrm{IR}_M) \qquad U_i^{\tilde{M}}(\theta_i) \geq U_i^M(\theta_i) \quad \forall i, \forall \theta_i.$$

Clearly, any manipulation of M that satisfies (IR_M) would also satisfy (IR) as long as M satisfies (IR). Hence, this requires (IR) to accommodate all acceptable collusive arrangements given passive beliefs, but it also includes arrangements supported by other, possibly extreme, beliefs. In particular, it means that the coalition can hold the members down to the same outside options, regardless of the principal's contract offer, thus limiting her ability to undermine collusion by manipulating their outside options. In fact, endowing the coalition with the ability to enforce any manipulation subject only to (IC) and (IR) is tantamount to assuming that the coalition enjoys the same commitment power as the principal. Although some may view this approach as assuming unrealistically powerful collusion, it can only strengthen our case if our implementation is robust against all such manipulations—a requirement we formalize as follows:

DEFINITION 1: A mechanism $M \in \mathcal{M}$ is *robustly collusion-proof* (RCP) if every $\tilde{M} \in \mathcal{M}_M$ gives the same expected payoff to the principal as mechanism M. A payoff $V \in \mathcal{V}$ is *RCP implementable* if there exists an RCP mechanism that implements V.

We next explore several features of our collusion-proof notion. Readers who wish to get to the main results may skip the remainder of this section.

5.2. *Implications and Comparison with Existing Notions*

Objective of the coalition

Our collusion-proofness notion imposes no restriction on the behavioral objective of the coalition. To see this, suppose, facing grand mechanism M, the coalition solves

$(\mathrm{C}_M(\alpha))$

$$\max_{M' \in \mathcal{M}_M} \mathbb{E}\left[\sum_{i\in N} \alpha_i(\tilde{\theta}) U_i^{M'}(\tilde{\theta}_i)\right]$$

for some $\alpha := (\alpha_1, \dots, \alpha_n) : \Theta \mapsto \mathbb{R}^n_+$. This formulation of the collusion problem encompasses a broad class of collusion possibilities, nesting many existing formulations as special cases. For instance, with $\alpha(\cdot) \equiv \mathbf{1}$, the objective function treats the agents rather symmetrically, as was assumed by LM. With $\alpha_i(\cdot) \equiv 1$ and $\alpha_j(\cdot) \equiv 0$ for all $j \neq i$, the representative caters to the interest of agent i at the expense of others, as will happen if agent i proposes a contract (see Mookherjee and Tsumagari (2004), for instance). In fact, any individually rational collusion agreement must correspond to some $\alpha(\cdot) \in \mathcal{A}$, where $\mathcal{A}$ is the set of all mappings $\alpha : \Theta \mapsto \mathbb{R}^n_+$. All these possible scenarios are captured in our notion: *If M is RCP, then $\forall \alpha \in \mathcal{A}$, every solution of* $(\mathrm{C}_M(\alpha))$ *gives the same expected payoff to the principal as mechanism M*. In fact, the principal need not even know the precise objective of the coalition.

Collusion prevention

Our collusion-proofness requirement does not rule out that collusion occurs on the equilibrium path, but rather ensures that the principal will not be harmed by collusion, even if it occurs. Clearly, this latter requirement is all that matters as far as the principal is concerned. Our requirement is, in fact, natural when the principal does not know the precise objective of the coalition (i.e., α). If the principal *does* know the objective, however, she can prevent collusion, given robust collusion-proofness.

PROPOSITION 1: *If a mechanism $M \in \mathcal{M}$ is RCP, then for each α with a non-empty solution to* ($\mathrm{C}_M(\alpha)$), *there exists a mechanism M_α that gives the same payoff as M to the principal and solves* ($\mathrm{C}_{M_\alpha}(\alpha)$).

PROOF: Suppose that $M = (q(\cdot), t(\cdot))$ is RCP and let $M_\alpha = (q_\alpha(\cdot), t_\alpha(\cdot))$ be a solution of ($\mathrm{C}_M(\alpha)$). Since M is RCP, M_α gives the same expected payoff to the principal as M. We prove that M_α solves ($\mathrm{C}_{M_\alpha}(\alpha)$). Since $M_\alpha \in \mathcal{RM}_M$, there exists a balanced-budget side contract $(\mu_\alpha(\cdot), y_\alpha(\cdot))$ such that $\forall\, \theta \in \Theta$,

$$v(q_\alpha(\theta)) = \mathbf{E}_{\mu_\alpha(\theta)}\big[v(q(\tilde{\theta}))\big] \quad \text{and} \quad t_\alpha(\theta) = \mathbf{E}_{\mu_\alpha(\theta)}[t(\tilde{\theta})] + y_\alpha(\theta). \tag{2}$$

Now pick any $\tilde{M} = (\tilde{q}(\cdot), \tilde{t}(\cdot)) \in \mathcal{RM}_{M_\alpha}$. Then there exists a balanced-budget side contract $(\mu(\cdot), y(\cdot))$ such that $\forall\, \theta$,

$$v(\tilde{q}(\theta)) = \mathbf{E}_{\mu(\theta)}\big[v(q_\alpha(\tilde{\theta}))\big] = \mathbf{E}_{\mu(\theta)}\big[\mathbf{E}_{\mu_\alpha(\tilde{\theta})}\big[v\big(q(\tilde{\tilde{\theta}})\big)\big]\big]$$

and

$$\tilde{t}(\theta) = \mathbf{E}_{\mu(\theta)}[t_\alpha(\tilde{\theta})] + y(\theta) = \mathbf{E}_{\mu(\theta)}\big[\mathbf{E}_{\mu_\alpha(\tilde{\theta})}\big[t(\tilde{\tilde{\theta}})\big]\big] + \mathbf{E}_{\mu(\theta)}[y_\alpha(\tilde{\theta})] + y(\theta),$$

where the last equalities follow from (2). Note that $\mathbf{E}_{\mu(\theta)}[\mathbf{E}_{\mu_\alpha(\tilde{\theta})}[\cdot]] = \mathbf{E}_{\bar{\mu}(\theta)}[\cdot]$ for some $\bar{\mu}(\theta) \in \Delta\Theta$ and if we let $\bar{y}(\theta) := \mathbf{E}_{\mu(\theta)}[y_\alpha(\tilde{\theta})] + y(\theta)$, then $\sum_{i\in N} \bar{y}_i(\theta) = \mathbf{E}_{\mu(\theta)}[\sum_{i\in N} y_{\alpha i}(\tilde{\theta})] + \sum_{i\in N} y_i(\theta) = 0$ for each $\theta \in \Theta$. Hence, $\tilde{M}$ is a reallocational manipulation of M or $\tilde{M} \in \mathcal{RM}_M$. We have thus shown that $\mathcal{RM}_{M_\alpha} \subset \mathcal{RM}_M$, which in turn implies $\mathcal{M}_{M_\alpha} \subset \mathcal{M}_M$. Since $M_\alpha \in \mathcal{M}_{M_\alpha}$, M_α must then solve ($\mathrm{C}_{M_\alpha}(\alpha)$). *Q.E.D.*

Relationship with other concepts

The most standard approach follows LM's *weak collusion-proofness*. This notion posits collusion organized by an uninformed third party who manipulates agents' reports in a way that is acceptable to the agents given their passive beliefs and maximizes their joint payoffs, but has no ability to reallocate their assignment. This notion can be formally stated in a way comparable to robust collusion-proofness. Given any grand mechanism M, say $\tilde{M} = (\tilde{q}, \tilde{t})$ is a *communicative manipulation* if there exists a balanced-budget side contract $(\mu, y) : \Theta \mapsto \Delta\Theta \times \mathbb{R}^n$ such that, for each $\theta \in \Theta$,

$$\tilde{t}(\theta) = \mathbf{E}_{\mu(\theta)}[t(\tilde{\theta})] + y(\theta) \quad \text{and} \quad \tilde{q}(\theta) = \mathbf{E}_{\mu(\theta)}[q(\tilde{\theta})]. \tag{3}$$

The second requirement shows inability to reallocate: the agents can influence the allocation only by manipulating their reports. Letting $\mathcal{CM}_M$ be the set of all communicative manipulations, we thus have $\mathcal{CM}_M \subset \mathcal{RM}_M$.

Formally, a mechanism M is *weakly collusion-proof* if it maximizes the objective of ($\mathrm{C}_M(\mathbf{1})$) among all communicative manipulations that satisfy (IC) and (IR_M). As LM show, when a mechanism is weakly collusion-proof, its outcome can be sustained in a collusive environment as a perfect Bayesian equilibrium. Our RCP notion encompasses this notion because we allow for any arbitrary α in the coalition's objective function and reallocational manipulations, and we assume (IR) instead of (IR_M) for collusive agreements.[12]

PROPOSITION 2: *If a mechanism M efficiently and RCP implements a payoff $V \in \mathcal{V}$, then it is weakly collusion-proof.*

PROOF: Suppose to the contrary that M is not weakly collusion-proof. Then there must be a mechanism $\tilde{M} \in \mathcal{CM}_M$ that satisfies (IC) and (IR_M), and generates a higher (expected) payoff for agents than M. Since $\mathcal{CM}_M \subset \mathcal{RM}_M$ and (IR_M) implies (IR), we have $\tilde{M} \in \mathcal{M}_M = \mathcal{RM}_M \cap \mathcal{M}$. Since M is RCP, $\tilde{M}$ yields the same payoff V to the principal. Consequently, $\tilde{M}$ must generate a strictly higher total surplus than M. However, this contradicts the fact that M efficiently implements V. *Q.E.D.*

Our notion does not subsume LM's *strong collusion-proofness*, which requires collusion-proofness relative to all possible out-of-equilibrium beliefs. Our notion allows for a range of reasonable out-of-equilibrium beliefs, including passive beliefs, but it implicitly rules out some extreme beliefs that are inconsistent with the agents' individual rationality.[13]

Our concept is also incompatible with a notion that requires agents' participation constraints to hold at the ex post, rather than interim, level. Ex post participation constraints are motivated either by agents' colluding on participation decisions (Dequiedt (2004)) or by their having an exit option from the grand mechanism ex post (Mookherjee and Tsumagari (2004)). These latter possibilities are not allowed in our model of collusion, so the ex post participation constraint is not required in our notion of collusion-proofness. Note that these other notions do not subsume our notion because we require robustness to many aspects of collusion discussed earlier.

[12]Jeon and Menicucci (2005) adopt the same notion except that they allow for reallocation by the agents. Hence, an RCP implementation will imply their notion as well.

[13]If such beliefs are admitted, the third party representative may be able to force a collusive proposal that may not guarantee a reservation utility for some agent. This will undermine implementation since the latter agent will refuse to participate in the grand contract.

6. RCP IMPLEMENTATION: UNCORRELATED TYPES

We now present our main collusion-proof implementation result. We begin with the case in which the types are uncorrelated. Except for type independence, we maintain the generality of the environments presented in Section 3.

THEOREM 1: *Suppose that types are uncorrelated. Then every $V \in \mathcal{V}$ is implementable by a robustly collusion-proof mechanism.*

PROOF: Fix any $V \in \mathcal{V}$ and suppose mechanism $M = (q, t) \in \mathcal{M}$ implements V. We now construct an RCP mechanism $\hat{M} = (\hat{q}, \hat{t}) \in \mathcal{M}$ that also implements V. Define $\hat{M} = (\hat{q}, \hat{t})$ such that $\hat{q}(\cdot) := q(\cdot)$ and that, for each $\theta \in \Theta$,

$$\begin{aligned} \hat{t}_i(\theta) := {} & \kappa_i v(q(\theta)) + \mathbb{E}_{\tilde{\theta}_{-i}}\big[t_i(\theta_i, \tilde{\theta}_{-i}) - \kappa_i v(q(\theta_i, \tilde{\theta}_{-i}))\big] \\ & - \frac{1}{n-1}\sum_{j \neq i} \mathbb{E}_{\tilde{\theta}_{-j}}\big[t_j(\theta_j, \tilde{\theta}_{-j}) - \kappa_j v(q(\theta_j, \tilde{\theta}_{-j}))\big] - \rho_i, \end{aligned} \tag{4}$$

where

$$\rho_i := \frac{1}{n-1}\mathbb{E}\left[(1-\kappa_i)v(q(\tilde{\theta})) - \sum_{j \neq i} t_j(\tilde{\theta})\right] \quad \text{and} \quad \sum_{i \in N} \kappa_i = 1.$$

Observe first that $\hat{t}(\cdot)$ gives the same interim transfers to the agents as $t(\cdot)$: $\forall i, \forall \theta_i \in \Theta_i$,

$$\begin{aligned} & \mathbb{E}_{\tilde{\theta}_{-i}}[\hat{t}_i(\theta_i, \tilde{\theta}_{-i})] \\ & \quad = \mathbb{E}_{\tilde{\theta}_{-i}}\big[\kappa_i v(q(\theta_i, \tilde{\theta}_{-i}))\big] + \mathbb{E}_{\tilde{\theta}_{-i}}\big[t_i(\theta_i, \tilde{\theta}_{-i}) - \kappa_i v(q(\theta_i, \tilde{\theta}_{-i}))\big] \\ & \qquad - \frac{1}{n-1}\sum_{j \neq i} \mathbb{E}_{\tilde{\theta}_j}\big[\mathbb{E}_{\tilde{\theta}_{-j}}\big[t_j(\tilde{\theta}_j, \tilde{\theta}_{-j}) - \kappa_j v(q(\tilde{\theta}_j, \tilde{\theta}_{-j}))\big]\big] - \rho_i \\ & \quad = \mathbb{E}_{\tilde{\theta}_{-i}}[t_i(\theta_i, \tilde{\theta}_{-i})]. \end{aligned} \tag{5}$$

It follows from (5) that $\hat{M}$ induces the same interim payoffs to the agents as M, so $\hat{M}$ satisfies (IC) and (IR). Furthermore, (5) means that the principal attains the same expected payoff from $\hat{M}$ as from M:

$$\mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i \in N} \hat{t}_i(\tilde{\theta})\right] = \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i \in N} t_i(\tilde{\theta})\right].$$

It now remains to show that $\hat{M}$ is RCP. To this end, we observe, $\forall \theta \in \Theta$,

$$\sum_{i \in N} \hat{t}_i(\theta) = v(q(\theta)) - \sum_{i \in N} \rho_i. \tag{6}$$

Consider any arbitrary reallocational manipulation of $\hat{M}$: $\tilde{M} = (\tilde{q}(\cdot), \tilde{t}(\cdot)) \in \mathcal{M}_{\hat{M}}$. Then there exists a balanced-budget side contract $(\mu, y) : \Theta \mapsto \Delta\Theta \times \mathbb{R}^n$ such that $\tilde{t}(\theta) = \mathbf{E}_{\mu(\theta)}[\hat{t}(\tilde{\theta})] + y(\theta)$ and $v(\tilde{q}(\theta)) = \mathbf{E}_{\mu(\theta)}[v(q(\tilde{\theta}))]$ for each $\theta \in \Theta$. Thus, for each $\theta \in \Theta$,

$$\begin{aligned}(7)\qquad \sum_{i\in N}\tilde{t}_i(\theta) &= \sum_{i\in N}\mathbf{E}_{\mu(\theta)}[\hat{t}_i(\tilde{\theta})] + \sum_{i\in N} y_i(\theta) = \mathbf{E}_{\mu(\theta)}\left[\sum_{i\in N}\hat{t}_i(\tilde{\theta})\right] \\ &= \mathbf{E}_{\mu(\theta)}\big[v(q(\tilde{\theta}))\big] - \sum_{i\in N}\rho_i = v(\tilde{q}(\theta)) - \sum_{i\in N}\rho_i,\end{aligned}$$

where the first and the last equalities follow from the definition of the reallocational manipulation, the second follows from the budget-balancedness of $y(\cdot)$, and the third follows from (6). It then follows from (7) that

$$(8)\qquad \mathbb{E}\left[v(\tilde{q}(\tilde{\theta})) - \sum_{i\in N}\tilde{t}_i(\tilde{\theta})\right] = \sum_{i\in N}\rho_i = \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i\in N}t_i(\tilde{\theta})\right],$$

thus proving that $\hat{M}$ is RCP. *Q.E.D.*

As seen from the proof, two features of our mechanism, $\hat{M}$, are central to its RCP implementation of an arbitrary mechanism M: First, as seen in (5), $\hat{M}$ preserves the same interim transfers as M, thus satisfying both (IR) and (IC) and giving the same expected payoff as M. Second, as seen in (6), the transfers $\hat{t}_i(\cdot)$ are aggregated so that the principal collects the ex post constant payoff equal to the expected payoff he would have enjoyed under M. This feature forces the coalition to become a "residual claimant" when it manipulates M, ensuring that the principal will attain the desired payoff regardless of how the agents behave, once they participate. The first feature, i.e., (IC) and (IR), guarantees an equilibrium in which the agents indeed participate in the mechanism.

Since every feasible payoff for the principal can be RCP implementable, the following result is immediate.

COROLLARY 1: *If $V^* \in \mathcal{V}$, then there exists an RCP mechanism that implements the noncollusive optimal payoff for the principal.*

The intuition behind our results can be made more transparent with the aid of Figure 1. Assume for a moment that there is no collusion problem. It is useful to think of the mechanism design problem as that of implementing a particular (expected) social surplus level, $\mathbb{E}[v(\tilde{q}(\theta)) + \sum_{i\in N} s_i(\tilde{q}(\theta), \theta)]$, i.e., the sum of all players' payoffs including that of the principal. The horizontal axis of Figure 1 depicts all implementable (expected) social surplus levels, with the

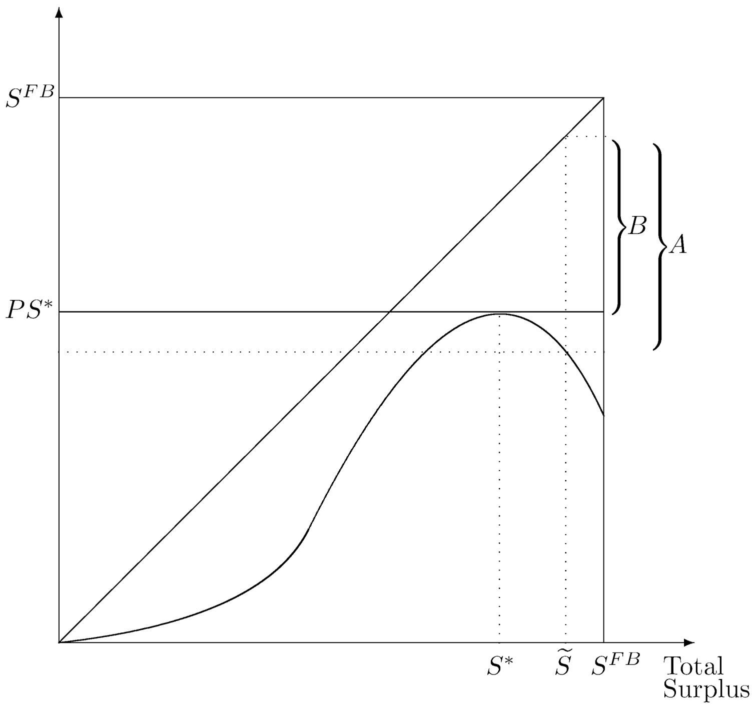

FIGURE 1.

highest level marking the first-best level, say.[14] Obviously, to achieve a given social surplus level, say $\tilde{S}$, requires a particular allocation rule $\tilde{q}(\cdot)$, and to implement the latter in turn requires giving away a certain amount (depicted by A in Figure 1) of information rent to the agents. Suppose the difference between the 45 degree line and the curve below it represents the minimal information rent that must be paid to the agents to implement the corresponding social surplus level. The curve below the 45 degree line then represents the (expected) surplus that accrues to the principal after paying off the rents to the agents. The figure depicts a common situation in which the principal's surplus is maximized at a below-first-best social surplus level, S^*, because of the rent-saving benefit gained from distorting the allocation.[15] In the absence of collusion, the principal would thus choose to implement S^*.

[14]In general, the first-best allocation may not be implementable even in the noncollusive environment. See Jehiel and Moldovanu (2001), for instance. In such a case, S^{FB} should be taken to mean the highest implementable surplus level.

[15]This situation is quite common in many mechanism design problems. For instance, an optimal auction often involves a binding reserve price or handicapping, both of which entail an inefficient allocation. Likewise, nonlinear pricing often induces too little consumption by the buyers.

We now introduce collusion and suppose that the principal proposes mechanism $\hat{M}$. Given mechanism $\hat{M}$, by inducing $(\tilde{q}(\cdot), \tilde{t}(\cdot))$ via manipulation say, the coalition members receive the joint payoff of

$$\sum_{i\in N}\big[\tilde{t}_i(\theta) + s_i(\tilde{q}(\theta), \theta)\big] = v(\tilde{q}(\theta)) + \sum_{i\in N} s_i(\tilde{q}(\theta), \theta) - \sum_{i\in N}\rho_i, \tag{9}$$

where the equality follows from (7). Hence, $\hat{M}$ forces the coalition to become residual claimants of the social surplus, after guaranteeing the principal an ex post constant surplus of

$$\sum_{i\in N}\rho_i = \mathbb{E}\left[v(q^*(\tilde{\theta})) - \sum_{i\in N} t_i^*(\tilde{\theta})\right],$$

which is described in the figure by the maximized level of the curve, PS^*. Hence, given $\hat{M}$, the coalition receives the difference between the 45 degree line and the horizontal line tangent at the principal's maximized surplus PS^*. Clearly, the mechanism does not eliminate the potential for coalitional manipulation, since the coalition now prefers S^{FB} over S^*, which the principal wishes to implement.

The reason the coalition cannot implement, via manipulation, S^{FB}—or any social surplus level other than S^* for that matter—is as follows: Since the coalition faces the same adverse selection problem that the principal faces without collusion, the amount of information rents the coalition must pay to the agents is described as before—by the difference between the 45 degree line and the curve. Under $\hat{M}$, the amount of surplus left to the coalition after paying off the principal is the gap between the 45 degree line and the horizontal tangent line, which is strictly less than the information rents needed to implement such a level, for any social surplus level that differs from S^*. If the collusion organizer wishes to implement $\tilde{S}$ (via manipulation), for instance, it will require the information rents of A, but the surplus left over after paying off the principal is only B, so implementing $\tilde{S}$ would entail a budget deficit of $A - B$ and is thus infeasible. Any deviation from S^* is not implementable for the same reason. This is possible precisely because S^* is optimal for the principal among all implementable social surplus levels.

REMARK 1: For an RCP implementation, the principal need not deal with the agents at all, opting rather to contract directly with their third-party representative. Fix an RCP mechanism $(\hat{t}, \hat{q})$ that implements any $V \in \mathcal{V}$. The principal can offer a menu $\{\hat{T}(\theta), \hat{q}(\theta)\}_{\theta\in\Theta}$ to the representative, where $\hat{T}(\cdot) := \sum_{i\in N}\hat{t}_i(\cdot)$ is a menu of total budgets. Facing such a contract, the representative will organize the coalition of agents and implement the desired payoff V for the principal. Absent collusion, such "delegation" would be trivial since the representative would act just like the original principal. The significance of the delegation is that it works even in the presence of collusion, i.e., even when the representative cares intrinsically about the welfare of the agents. In fact, such delegation may provide a practical way to implement an RCP mechanism. For instance, instead of hiring and supervising individual suppliers, a buyer may wish to outsource the job to an intermediary (prime contractor) who organizes and supervises a network of suppliers (subcontractors).

REMARK 2: The design of transfer rules in (4) is reminiscent of the Arrow–d'Aspremont–Gérard-Varet mechanism (Arrow (1979) and d'Aspremont and Gérard-Varet (1979)). Their transfers preserve the incentives of the Vickrey–Clarke–Groves mechanism with zero aggregate transfer. Ours implements the interim payoffs of any original mechanism with an ex post constant "surplus" for the principal. Esö and Futos (1999) suggested a similar transfer rule that gives rise to ex post constant "revenue" as an optimal mechanism for a risk-averse seller in a single-unit auction. See also Bose, Ozdenoren, and Pape (2005). These mechanisms serve much different purposes in their analysis. Nonetheless, our model is more general and our mechanism subsumes theirs as a special case (with $v(\cdot)$ being constant). For this reason, our construction would generalize their results. For instance, our RCP mechanism for implementing V^* would be (noncollusive) optimal even if the principal were risk averse.[16]

7. RCP IMPLEMENTATION: CORRELATED TYPES

We now turn to the case in which the agents' types are correlated. In this case, we already know from LM (2000) that collusion cannot be prevented for free in a public good model with two agents and two types. As we show below, however, our collusion-proof implementation result holds even in a large class of correlated type environments if there are more than two agents. Given the well-known result by Crémer and McLean (1985, 1988), this implies that the principal can extract full rents and implement the first-best outcome *even if collusion is possible*.

Consider our general environment in Section 3, but assume that the joint type space Θ is finite. (The finite type space will enable us to utilize linear algebra, as will be seen.) Specifically, we assume that the support of agent $i \in N$'s type is given by $\Theta_i = \{\theta_i^1, \dots, \theta_i^{\ell_i}\}$ with $\ell_i = |\Theta_i| \geq 2$. Let $L := \prod_{i\in N} \ell_i$. It is useful to index all elements of Θ (i.e., all possible type profiles) in an arbitrary order so that $\Theta = \{\theta^1, \dots, \theta^L\}$. We then suppose that each type profile $\theta \in \Theta$ is

[16]More precisely, the RCP mechanism that implements V^* is the unique optimal mechanism for a principal with a strictly concave von Neumann–Morgenstern utility function $u_0(a(x) - \sum_{i\in N} t_i)$.

realized by a joint probability $\mu^0(\theta) \in [0,1]$ such that $\sum_{\theta' \in \Theta} \mu^0(\theta') = 1$, where $\mu^0 := (\mu^0(\theta^1), \dots, \mu^0(\theta^L))'$ represents the vector of joint probabilities of all type profiles listed in the order mentioned above.

Fix any mechanism $M = (q(\cdot), t(\cdot)) \in \mathcal{M}$, attainable in a noncollusive environment. As before, we consider a new mechanism, $\hat{M} = (q(\cdot), \hat{t}(\cdot))$, that satisfies two properties on the transfer rule: (a) it satisfies both (IC) and (IR) and yields the same interim transfers to the agents as M, and (b) it ensures that the principal enjoys an ex post constant surplus that equals the expected surplus she would enjoy under M.

Formally, (a) holds if

$$\sum_{\theta_{-i} \in \Theta_{-i}} \mu^0(\theta_i^k, \theta_{-i}) \hat{t}_i(\theta_i^{k'}, \theta_{-i}) = \sum_{\theta_{-i} \in \Theta_{-i}} \mu^0(\theta_i^k, \theta_{-i}) t_i(\theta_i^{k'}, \theta_{-i}) \quad \forall\, i \text{ and } \forall\, \theta_i^k, \theta_i^{k'} \in \Theta_i \tag{10}$$

and (b) holds if

$$\sum_{i \in N} \hat{t}_i(\theta') = v(q(\theta')) - \rho \quad \forall\, \theta' \in \Theta, \tag{11}$$

where $\rho := \mathbb{E}[v(q(\tilde{\theta})) - \sum_{i \in N} t_i(\tilde{\theta})]$. Since t satisfies (IC) and (IR), (10) ensures that $\hat{t}$ satisfies (IC) and (IR) and yields the same interim payoffs to all the players as M. Meanwhile, (11) makes the agents residual claimants. Together, these two features guarantee that $\hat{M}$ implements the optimal noncollusive mechanism M in a collusion-proof fashion.

We describe these two restrictions by a system of linear equations. To begin, define for each i and $\theta_i^k, \theta_i^{k'} \in \Theta_i$ with $\theta_i^{k'} \neq \theta_i^k$,

$$T_i(\theta_i^k) := \sum_{\theta_{-i} \in \Theta_{-i}} \mu^0(\theta_i^k, \theta_{-i}) t_i(\theta_i^k, \theta_{-i}),$$

$$S_i(\theta_i^k, \theta_i^{k'}) := \sum_{\theta_{-i} \in \Theta_{-i}} \mu^0(\theta_i^k, \theta_{-i}) t_i(\theta_i^{k'}, \theta_{-i}).$$

In words, $T_i(\theta_i^k)$ and $S_i(\theta_i^k, \theta_i^{k'})$ denote the interim transfer agent i of type θ_i^k receives when reporting truthfully and when misreporting to be of type $\theta_i^{k'}$, respectively, given that all other agents report truthfully. We can then form interim transfer vectors as

$$T_i := (T_i(\theta_i^k))_{\theta_i^k \in \Theta_i} \quad \text{and} \quad S_i := \big(S_i(\theta_i^k, \theta_i^{k'})\big)_{\theta_i^k, \theta_i^{k'} \in \Theta_i, \theta_i^k \neq \theta_i^{k'}}.$$

We next form a vector of transfers for the new mechanism. For each $i \in N$, let

$$\hat{t}_i := (\hat{t}_i(\theta^1), \dots, \hat{t}_i(\theta^L))'.$$

(That is, the elements of the vector are listed in the order of $\Theta = \{\theta^1, \dots, \theta^L\}$.) Next, we form a matrix P_i, of size $\ell_i \times L$, which represents the probabilities over the reported types of all agents when agent i reports truthfully. Specifically, the mth element of a row that corresponds to $T_i(\theta_i^k)$ has probability $\mu^0(\theta_i^k, \theta_{-i})$ if $\theta^m = (\theta_i^k, \theta_{-i})$ and zero otherwise. Similarly, a matrix B_i, of size $\ell_i(\ell_i - 1) \times L$, represents the probabilities over the reported types when agent i lies. Specifically, the mth element of a row that corresponds to $S_i(\theta_i^k, \theta_i^{k'})$ has probabilities $\mu^0(\theta_i^k, \theta_{-i})$ if $\theta^m = (\theta_i^{k'}, \theta_{-i})$ and zero otherwise.[17] Then (10) is expressed as

$$\begin{pmatrix} P_i \\ B_i \end{pmatrix} \times \hat{t}_i = \begin{pmatrix} T_i \\ S_i \end{pmatrix} \quad \forall i \in N.$$

To combine with the second property, we define a vector of length L:

$$v - \rho := \big(v(q(\theta^1)) - \rho, \dots, v(q(\theta^L)) - \rho\big)'.$$

Then (10) and (11) are described in matrix forms as

$$\text{(12)} \qquad \begin{pmatrix} \Pi_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Pi_n \\ I_L & \cdots & I_L \end{pmatrix} \times \begin{pmatrix} \hat{t}_1 \\ \vdots \\ \hat{t}_n \end{pmatrix} = \begin{pmatrix} T_1 \\ S_1 \\ \vdots \\ T_n \\ S_n \\ v - \rho \end{pmatrix},$$

where $\Pi_i := \binom{P_i}{B_i}$ and I_L is the identity matrix of size L.

The next condition proves to be sufficient for the existence of a solution to (12):

[17]It is useful to consider a specific example. Suppose there are three agents and each has two types, $\Theta_i = \{1, 2\}$, $i = 1, 2, 3$. Suppose Θ is indexed as $\{111, 112, 121, 122, 211, 212, 221, 222\}$, where the type profile ijk refers to agent 1 being of type i, agent 2 being of type j, and agent 3 being of type k. Then

$$P_2 = \begin{pmatrix} \mu^0(111) & \mu^0(112) & 0 & 0 & \mu^0(211) & \mu^0(212) & 0 & 0 \\ 0 & 0 & \mu^0(121) & \mu^0(122) & 0 & 0 & \mu^0(221) & \mu^0(222) \end{pmatrix}$$

and

$$B_2 = \begin{pmatrix} 0 & 0 & \mu^0(111) & \mu^0(112) & 0 & 0 & \mu^0(211) & \mu^0(212) \\ \mu^0(121) & \mu^0(122) & 0 & 0 & \mu^0(221) & \mu^0(222) & 0 & 0 \end{pmatrix},$$

where $\mu^0(ijk)$ is the probability of agents 1, 2, and 3 being, respectively, of types i, j, and k. The first and second rows of the above P_2 correspond to $T_2(1)$ and $T_2(2)$, respectively. The first and second rows of B_2 correspond to $S_2(1, 2)$ and $S_2(2, 1)$, respectively.

CONDITION (PI′): *There exist agents i and j such that*

$$\text{rank}\begin{pmatrix}\Pi_i \\ \Pi_j\end{pmatrix} = \text{rank}(\Pi_i) + \text{rank}(\Pi_j) - 1. \tag{13}$$

Condition (PI′) requires that the spaces spanned by the rows of Π_i and those spanned by the rows of Π_j should not intersect except for a one-dimensional vector space, which accounts for a redundancy in the system of equations.[18] Intuitively, it requires that the untruthful reports by agents i and all possible reports by agent j induce distinct probability distributions over the entire report profiles, assuming all others report truthfully. This feature provides the flexibility needed to mimic the incentive design of t and at the same time makes the agents residual claimants.[19]

Our main results follow.

LEMMA 1: *Given Condition* (PI′), *a solution*, $\hat{t} \in \mathbb{R}^{nL}$, *to* (12) *exists*.

See Appendix A for the proof.

THEOREM 2: *Given Condition* (PI′), *every* $V \in \mathcal{V}$ *is RCP implementable*.

[18]The redundancy comes from an accounting identity associated with the fact that the equilibrium transfers specified in top and bottom parts of the system must be consistent with each other. To be specific, if we premultiply the equations of (12) that correspond to T_i by the probability vector $\mu^{0\prime}$, we obtain the aggregate expected transfers in equilibrium:

$$\sum_{i\in N}\mu^{0\prime}\cdot\hat{t}_i = \sum_{i\in N}\mathbb{E}[\hat{t}_i(\tilde{\theta})] = \text{LHS} = \text{RHS} = \sum_{i\in N}\sum_{\theta_i\in\Theta_i}T_i(\theta_i) = \sum_{i\in N}\mathbb{E}[t_i(\tilde{\theta})].$$

This must be consistent with the restrictions on the aggregate transfers in expected value. In particular, we obtain the same equation by premultiplying the bottom L equations of (12) with the probability vector $\mu^{0\prime}$:

$$\begin{aligned}\sum_{i\in N}\mathbb{E}[\hat{t}_i(\tilde{\theta})] &= \text{LHS} = \text{RHS} = \mu^{0\prime}\cdot(v-\rho)\\ &= \mathbb{E}\big[v(q(\tilde{\theta}))\big] - \mathbb{E}\Big[v(q(\tilde{\theta})) - \sum_{i\in N}t_i(\theta)\Big] = \sum_{i\in N}\mathbb{E}[t_i(\tilde{\theta})].\end{aligned}$$

[19]Similar conditions have appeared in the literature in the context of repeated games and static mechanism design with budget balancing (see Fudenberg, Levine, and Maskin (1994, 1995) and Kosenok and Severinov (2004)). *Pairwise identifiability for a pair of agents i and j*, considered by Fudenberg, Levine, and Maskin, requires the same rank condition, except that Π_i represents probability distributions that correspond to different *strategies* rather than all pairs of reports and states. So Π_i has $\ell_i^{\ell_i}$ rows in the pairwise identifiability, whereas Π_i in our Condition (PI′) has ℓ_i^2 rows. In fact, it can be shown that the pairwise identifiability is weaker than Condition (PI′) in the sense that if (13) holds for a pair of agents i and j, then pairwise identifiability holds for the same pair. Several conditions proposed by Kosnok and Severinov (2004) are closely related to Condition (PI′). In fact, the genericity of Condition (PI′) follows from the genericity of one of their conditions, as stated in Lemma 2.

PROOF: Fix any $M = (q, t) \in \mathcal{M}$. Then, given Condition (PI′) and Lemma 1, we can consider a mechanism $\hat{M} = (q(\cdot), \hat{t}(\cdot))$, where $\hat{t}(\cdot)$ solves the system of linear equations in (12). Since $\hat{t}(\cdot)$ satisfies (10), the interim transfers and, thus, the interim payoffs for the agents are precisely the same under $\hat{M}$ as under M for any possible report each agent may make, assuming that all other agents report truthfully. This guarantees that $\hat{M}$ satisfies (IC) and (IR) and yields the same interim payoffs to all players as M. Next, since $\hat{t}$ satisfies (11), the same argument as in the proof of Theorem 1 proves that $\hat{M}$ is RCP. *Q.E.D.*

It can be readily checked that Condition (PI′) fails generically if there are only two agents. In that case, generically, the right-hand side of (13) becomes $\ell_i\ell_j + \min\{\ell_i^2, \ell_j^2\} - 1$, thus exceeding the rank of the stacked matrix on its left-hand side, which equals $\ell_i\ell_j$ generically.[20] This is consistent with LM (2000), which finds that the principal is strictly worse off from collusion and that the principal's optimized payoff is continuous at zero correlation if the agents' types are correlated. The latter finding of LM contrasts with the noncollusive mechanism design, which displays a discontinuous shift from a typical second-best outcome to a full-extraction first-best outcome when an arbitrarily small amount of type correlation is added.

If there are more than two agents, however, Condition (PI′) holds quite generally, so LM's (2000) result does not hold.[21] The following result due to Kosenok and Severinov (2004) makes the statement precise.

LEMMA 2—Kosenok and Severinov: *Suppose that $n \geq 3$ and that, if $n = 3$, at least one agent has more than two types. Then Condition (PI′) holds for generic $\mu^0(\cdot)$.*[22]

PROOF: This result follows from steps 3–5 in the proof of Lemma 3 of Kosenok and Severinov (2004, pp. 26–28). In particular, they prove that, given the condition, the matrix on the left-hand side of (13) has a one-dimensional kernel for a generic $\mu^0(\cdot)$ for $j := \arg\min_{k\in N} \ell_k$ and $i := \arg\min_{k\in N\setminus\{j\}} \ell_k$. Hence, (13) holds generically for that pair. *Q.E.D.*

Given Lemma 2, our collusion-proof implementation holds generically for any $n \geq 3$, with the additional requirement that an agent must have more than

[20]Generically, each matrix on the right-hand side has a full rank since different reports by an agent induce different distributions over the other agent's reports when the latter reports truthfully. Hence, the ranks sum to $(\ell_i + \ell_j)\min\{\ell_i, \ell_j\} - 1 = \ell_i\ell_j + \min\{\ell_i^2, \ell_j^2\} - 1$.

[21]Laffont and Martimort's result with two agents means, however, that collusion will matter again if a subgroup of two agents is collusive. As will be noted in Remark 3, our method does not generalize to the subgroup collusion problem if types are correlated. In this sense, LM's point remains valid.

[22]This condition means that the probability distributions μ^0 for which the condition holds have full Lebesque measure in the $(n-1)$-dimensional simplex.

two types if $n = 3$. It is not difficult to see why adding a new agent makes it easier to satisfy the properties required for collusion-proofness. Suppose there are two agents with two possible types. Then the transfer rule specifies eight transfer amounts: A transfer amount is specified for each of four joint realizations of types for each of the two agents. Meanwhile, the number of equations required by (12) is 12, so the system has no solution. Intuitively, the transfer rule does not give a sufficient number of degrees of freedom to "sell the firm to the agents" while preserving the original incentive design of t. As the number of agents increases, the number of unknowns (the transfer amounts to be specified) increases multiplicatively while the number of restrictions increases only additively. The reason is that the restrictions implied by (a) need to hold only at the interim level. To be concrete, suppose that there is another agent with three types.[23] Then, the number of transfer amounts to be specified grows to 36 (i.e., a transfer specified for each of 12 joint type realizations for each of the three agents), whereas the number of equations required by (12) grows only to 29. This creates enough flexibility to satisfy both (a) and (b).[24]

As is well known from Crémer and McLean (1988), the full-extraction first-best outcome is implementable for generic $\mu^0(\cdot)$.[25] Lemma 2 implies that the outcome is attainable even in the collusive environment in a broad set of circumstances.

COROLLARY 2: *Given the condition of Lemma 2, for a generic $\mu^0(\cdot)$, there exists an RCP mechanism that implements the full-extraction first-best outcome.*

8. COLLUSION BY A SUBGROUP OF AGENTS

So far, we have only considered the possibility of collusion involving all agents. In many situations, however, only a subset of agents may be in a position to collude. For instance, in construction procurement auctions in which

[23] If the third agent has two types, then the system in (12) has more unknowns than the number of equations, but the matrix on its left-hand side has a two-dimensional kernel, whereas the system has only one-dimensional redundancy. So, the solution does not exist generically.

[24] As footnote 23 indicates, the comparison between the number of equations and the number of unknowns does not inform us of the existence of a solution because of possible linear dependence across the required equations. Nonetheless, the comparison is suggestive of how the properties required for collusion-proofness can be met.

[25] The full-extraction first-best outcome is defined as in Crémer and McLean (1988). In our context, it means that (IR) is binding for all agents for all types and the implemented allocation rule has

$$q^*(\theta) \in \arg\max_{q \in \mathcal{Q}} v(q) + \sum_{i \in N} s_i(q(\theta), \theta),$$

for all $\theta \in \Theta$. The conditions for this outcome to be implementable are shown to be generic.

both local and nonlocal firms compete, local firms may be able to collude more effectively, based on their regular contacts and trade association relationships. Can such collusion be prevented?

Preventing collusion by a subgroup of agents introduces a new consideration in mechanism design, since one has to consider the effect of collusion on the incentives of noncollusive agents. In particular, a coalition may gain from preying on the noncollusive agents by shifting rents away from them.[26] Hence, an important element of a collusion-proof mechanism is to protect noncollusive agents' interests/incentives appropriately. This section discusses how the basic idea of collusion-proof implementation generalizes to this case. We again restrict the economic environment to uncorrelated types; a remark will be made later on the correlated types case.

We begin with the model of subgroup collusion. To this end, consider a coalition $C \subset N$ of agents with $1 < |C| \le n$. The time line of the game and its modeling framework are analogous to the case with collusion by the grand coalition. Hence, the robust collusion-proofness concept generalizes naturally to this partial collusion case. To begin, we define a side contract among coalition C, called a *C-side contract*, by a pair of functions $(\mu^C, y^C): \Theta_C \mapsto \Delta\Theta_C \times \mathbb{R}^{|C|}$, where μ^C determines the probability distribution of reports on the coalition's types in Θ_C. Then, for any direct mechanism $M = (q(\cdot), t(\cdot))$, we call $\tilde{M} = (\tilde{q}(\cdot), \tilde{t}(\cdot)) \in \mathcal{M}$ its *reallocational manipulation by C* if there exists a balanced-budget C-side contract $(\mu^C, y^C): \Theta_C \mapsto \Delta\Theta_C \times \mathbb{R}^{|C|}$, such that $\forall\, \theta$,

$$\tilde{t}_i(\theta) = \begin{cases} \mathbf{E}_{\mu^C(\theta_C)}[t_i(\theta_{N\setminus C}, \tilde{\theta}_C)] + y_i^C(\theta_C), & \text{if } i \in C, \\ \mathbf{E}_{\mu^C(\theta_C)}[t_i(\theta_{N\setminus C}, \tilde{\theta}_C)], & \text{if } i \in N\setminus C, \end{cases} \tag{14}$$

$$v(\tilde{q}(\theta)) = \mathbf{E}_{\mu^C(\theta_C)}\bigl[v(q(\theta_{N\setminus C}, \tilde{\theta}_C))\bigr], \tag{15}$$

and

$$s_i(\tilde{q}(\theta), \theta) = \mathbf{E}_{\mu^C(\theta_C)}\bigl[s_i(q(\theta_{N\setminus C}, \tilde{\theta}_C), \theta)\bigr] \quad \forall\, i \in N\setminus C. \tag{16}$$

Note that a reallocational manipulation is required to be undetectable not only to the principal (see (15)), but also to the noncollusive agents (see (16)). In a single-unit auction with two collusive bidders and one noncollusive bidder, for instance, the latter restriction means that reallocation of the object between the two colluders is possible only when the object is initially assigned to one of the two collusive bidders.[27]

[26]Whether this problem arises depends on the grand mechanism in place. For instance, if a subset of bidders collude in a first-price auction, this may actually benefit noncollusive bidders.

[27]Consider, for instance, a single-unit interdependent value auction in which a bidder $i = 1, 2, 3$ realizes the valuation of $a_i(\theta_1, \theta_2, \theta_3)$ from winning the good (and zero for not winning), and

Similar to the grand coalition case, we say that a reallocational manipulation of M by coalition C, $\tilde{M}$ is *feasible* if it satisfies

$$(\text{IR}^C) \qquad U_i^{\tilde{M}}(\theta_i) \geq \overline{U}_i(\theta_i) \quad \forall i \in C, \theta_i$$

and

$$(\text{IC}^C) \qquad U_i^{\tilde{M}}(\theta_i) \geq u_i^{\tilde{M}}(\tilde{\theta}_i, \theta_i) \quad \forall i \in C, \theta_i, \tilde{\theta}_i,$$

and we let $\mathcal{M}_M^C$ denote the set of all feasible reallocational manipulations of M by coalition C. Note that a feasible reallocational manipulation by the coalition need not satisfy incentive compatibility and individual rationality of the agents outside that coalition, since the latter does not care about the noncollusive agents. Instead, we impose these conditions as part of the coalition-proofness requirements.

DEFINITION 2: A direct mechanism M is *robustly collusion-proof* (RCP) *with respect to coalition* $C \subset N$ if $\mathcal{M}_M^C \subset \mathcal{M}$ and every manipulation in $\mathcal{M}_M^C$ gives the same expected payoff to the principal as does mechanism M.

The notion of collusion-proofness here is essentially the same as before, except for the additional requirement that every reallocational manipulation by a coalition must be also incentive compatible and individually rational to noncollusive agents.[28] The extra requirement is added to protect the interests/incentives of the noncollusive agents against possible manipulation by the coalition, thus ensuring their participation and ultimately the intended payoff of the principal. Suppose the subcoalition wishes to induce a manipulation $\tilde{M} = (\tilde{q}, \tilde{t}) \in \mathcal{M}_M^C$ that violates either incentive compatibility or individual rationality of some noncollusive agent. If such a manipulation is anticipated, then the latter agent may lie about his type or not participate in M, in which case the principal may not receive the same expected payoff as M.

---

suppose for simplicity that the seller never retains the good. Then the reallocation ability by a coalition $C = \{1, 2\}$ means that a resulting allocation $\tilde{q}$ must satisfy, $\forall \theta = (\theta_1, \theta_2, \theta_3)$,

$$\tilde{q}_1(\theta) + \tilde{q}_2(\theta) = q_1(\tilde{\theta}_1, \tilde{\theta}_2, \theta_3) + q_2(\tilde{\theta}_1, \tilde{\theta}_2, \theta_3)$$

for any manipulation $(\tilde{\theta}_1, \tilde{\theta}_2)$ by C. It then follows that

$$\begin{aligned} s_3(\tilde{q}(\theta), \theta) &= (1 - \tilde{q}_1(\theta) - \tilde{q}_2(\theta))a_3(\theta) \\ &= \big(1 - q_1(\tilde{\theta}_1, \tilde{\theta}_2, \theta_3) - q_2(\tilde{\theta}_1, \tilde{\theta}_2, \theta_3)\big)a_3(\theta) = s_3(q(\tilde{\theta}_1, \tilde{\theta}_2, \theta_3), \theta), \end{aligned}$$

which implies (16).

[28]This requirement is superfluous in the case of the grand coalition since a feasible manipulation is assumed to satisfy incentive compatibility and individual rationality for all agents.

The additional requirement in collusion-proofness translates into an additional property to be satisfied in mechanism design. We say a mechanism $M = (q(\cdot), t(\cdot))$ is *ex post implementable* if it is *ex post individually rational*,

$$s_i(q(\theta_i, \theta_{-i}), \theta) - t_i(\theta_i, \theta_{-i}) \geq \overline{U}_i(\theta_i) \quad \forall\, i, \forall\, \theta_i, \forall\, \theta_{-i}, \tag{17}$$

and *ex post incentive compatible*,

$$\begin{aligned} &s_i(q(\theta_i, \theta_{-i}), \theta) - t_i(\theta_i, \theta_{-i}) \\ &\quad \geq s_i(q(\theta_i', \theta_{-i}), \theta) - t_i(\theta_i', , \theta_{-i}) \quad \forall\, i, \forall\, \theta_i, \forall\, \theta_i', \forall\, \theta_{-i}. \end{aligned} \tag{18}$$

Let $\mathcal{V}^{\mathrm{EP}}$ denote the set of payoffs that the principal can attain by ex post implementable mechanisms. Clearly, $\mathcal{V}^{\mathrm{EP}} \subset \mathcal{V}$. Later, we provide a clearer sense about $\mathcal{V}^{\mathrm{EP}}$ by presenting a sufficient condition for ex post implementability.

THEOREM 3: *Suppose the types are uncorrelated and fix any two agents $i, j \in N$. Then any $V \in \mathcal{V}^{\mathrm{EP}}$ is implementable by a mechanism that is RCP with respect to any coalition C that contains $\{i, j\}$.*

PROOF: Fix any $V \in \mathcal{V}^{\mathrm{EP}}$ and suppose $M = (q, t) \in \mathcal{M}$ ex post implements V. For any two agents $i, j \in N$, we construct a new mechanism $\bar{M}_{ij} \in \mathcal{M}$ that would RCP implement V. Let a mechanism $\bar{M}_{ij} := (q(\cdot), \bar{t}(\cdot))$ be such that, for each $k \neq i, j$, $\forall\, \theta$,

$$\bar{t}_k(\theta) = t_k(\theta)$$

and, for i, $\forall\, \theta \in \Theta$,

$$\begin{aligned} \bar{t}_i(\theta) = {} & \frac{1}{2}\left[v(q(\theta)) - \sum_{k \neq i,j} t_k(\theta)\right] \\ & + \mathbb{E}_{\tilde{\theta}_{-i}}\left[t_i(\theta_i, \tilde{\theta}_{-i}) - \frac{1}{2}\left\{v(q(\theta_i, \tilde{\theta}_{-i})) - \sum_{k \neq i,j} t_l(\theta_i, \tilde{\theta}_{-i})\right\}\right] \\ & - \mathbb{E}_{\tilde{\theta}_{-j}}\left[t_j(\theta_j, \tilde{\theta}_{-j}) - \frac{1}{2}\left\{v(q(\theta_j, \tilde{\theta}_{-j})) - \sum_{k \neq i,j} t_k(\theta_j, \tilde{\theta}_{-j})\right\}\right] \\ & - \mathbb{E}\left[\frac{1}{2}\left\{v(q(\tilde{\theta})) - \sum_{k \neq i,j} t_k(\tilde{\theta})\right\} - t_j(\tilde{\theta})\right], \end{aligned}$$

and symmetrically for j; i.e., $\bar{t}_j$ defined exactly the same with the roles of i and j switched.

Observe first that, $\forall\, k$, $\forall\, \theta_k \in \Theta_k$,

$$\mathbb{E}_{\tilde{\theta}_{-k}}[\bar{t}_k(\theta_k, \tilde{\theta}_{-k})] = \mathbb{E}_{\tilde{\theta}_{-k}}[t_k(\theta_k, \tilde{\theta}_{-k})], \tag{19}$$

so $\bar{M}$ satisfies (IC) and (IR) and attains the same value as M. Hence, $\bar{M}$ implements V.

We next show that $\bar{M}$ is RCP with respect to any coalition C that contains agents i and j. To show this, fix any such coalition C and choose any feasible reallocational manipulation of $\bar{M}$ by C, say $\tilde{M} = (\tilde{q}(\cdot), \tilde{t}(\cdot)) \in \mathcal{M}^C_{\bar{M}}$. Then there exists a balanced-budget C-side contract $(\mu^C(\cdot), y^C(\cdot))$ that satisfies (14), (15), and (16). Furthermore, $\tilde{M}$ satisfies (IC^C) and (IR^C), so it is incentive compatible and individually rational for any collusive agent in C. Consider now any $k \in N\backslash C$. Then, $\forall\, \theta_k, \theta'_k \in \Theta_k$, $\forall\, \theta_{-k} \in \Theta_{-k}$,

$$
\begin{aligned}
&s_k(\tilde{q}(\theta_k, \theta_{-k}), \theta) + \tilde{t}_k(\theta_k, \theta_{-k}) \\
&\quad = \mathbf{E}_{\mu^C(\theta_C)}\big[s_k(q(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C), \theta) + \bar{t}_k(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C)\big] \\
&\quad = \mathbf{E}_{\mu^C(\theta_C)}\big[s_k(q(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C), \theta) + t_k(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C)\big] \\
&\quad \geq \mathbf{E}_{\mu^C(\theta_C)}\big[s_k(q(\theta'_k, \theta_{N\backslash C-k}, \tilde{\theta}_C), \theta) + t_k(\theta'_k, \theta_{N\backslash C-k}, \tilde{\theta}_C)\big] \\
&\quad = s_k(\tilde{q}(\theta'_k, \theta_{-k}), \theta) + \tilde{t}_k(\theta'_k, \theta_{-k}),
\end{aligned}
$$

where the first and last equalities follow from (14) and (16), the second follows from the construction of $\bar{t}_k$ for $k \in N\backslash C$, and the lone inequality follows from ex post implementability of M. Likewise, $\forall\, k \in N\backslash C$, $\forall\, \theta_k, \theta_{-k}$,

$$
\begin{aligned}
&s_k(\tilde{q}(\theta), \theta) + \tilde{t}_k(\theta) \\
&\quad = \mathbf{E}_{\mu^C(\theta_C)}\big[s_k(q(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C), \theta) + t_k(\theta_k, \theta_{N\backslash C-k}, \tilde{\theta}_C)\big] \\
&\quad \geq \mathbf{E}_{\mu^C(\theta_C)}[\overline{U}_k(\theta_k)] = \overline{U}_k(\theta_k).
\end{aligned}
$$

These inequalities prove that $\tilde{M}$ is also incentive compatible and individually rational for agent $k \in N\backslash C$. In sum, $\tilde{M}$ must satisfy (IC) and (IR). Since $\tilde{M}$ is an arbitrary element of $\mathcal{M}^C_{\bar{M}}$, this proves that $\mathcal{M}^C_{\bar{M}} \subset \mathcal{M}$.

It now remains to show that $\tilde{M}$ implements V. Observe, $\forall\, \theta \in \Theta$,

$$
\begin{aligned}
\sum_{k\in N} \tilde{t}_k(\theta) &= \sum_{k\in N} \mathbf{E}_{\mu^C(\theta_C)}[\bar{t}_k(\theta_{N\backslash C}, \tilde{\theta}_C)] + \sum_{k\in C} y^C_k(\theta_C) \\
&= \sum_{k=i,j} \mathbf{E}_{\mu^C(\theta_C)}[\bar{t}_k(\theta_{N\backslash C}, \tilde{\theta}_C)] + \sum_{k\neq i,j} \mathbf{E}_{\mu^C(\theta_C)}[\bar{t}_k(\theta_{N\backslash C}, \tilde{\theta}_C)] \\
&= \mathbf{E}_{\mu^C(\theta_C)}\Bigg[\sum_{k=i,j} \bar{t}_k(\theta_{N\backslash C}, \tilde{\theta}_C)\Bigg] + \mathbf{E}_{\mu^C(\theta_C)}\Bigg[\sum_{k\neq i,j} t_k(\theta_{N\backslash C}, \tilde{\theta}_C)\Bigg] \\
&= \mathbf{E}_{\mu^C(\theta_C)}\Bigg[v(q(\theta_{N\backslash C}, \tilde{\theta}_C)) - \sum_{k\neq i,j} t_k(\theta_{N\backslash C}, \tilde{\theta}_C)\Bigg]
\end{aligned}
$$

$$- \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{k\in N} t_k(\tilde{\theta})\right] + \mathbf{E}_{\mu^C(\theta_C)}\left[\sum_{k\neq i,j} t_k(\theta_{N\setminus C}, \tilde{\theta}_C)\right]$$

$$= v(\tilde{q}(\theta)) - \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{k\in N} t_k(\tilde{\theta})\right],$$

where the first equality follows from (14), the second follows from the budget-balancedness of the side contract, the third follows from (14) and the switching of expectation and summation, the fourth follows from the construction of $\bar{t}(\cdot)$, and the fifth follows from collecting terms and from (15). It follows that

$$\mathbb{E}\left[v(\tilde{q}(\tilde{\theta})) - \sum_{i\in N} \tilde{t}_i(\tilde{\theta})\right] = \mathbb{E}\left[v(q(\tilde{\theta})) - \sum_{i\in N} t_i(\tilde{\theta})\right] = V, \tag{20}$$

proving that $\tilde{M}$ implements V. We thus conclude that $\bar{M}$ is RCP with respect to C. *Q.E.D.*

According to this proposition, the principal needs to identify *only* two members of any possible coalition to handle *any* collusion that involves the two, including the grand collusion. Hence, neither the principal nor any noncollusive agents need to know the precise size or the membership of the coalition, as long as two core members of collusion are identified. The RCP mechanism that does this has three features: (i) As before, the mechanism involves selling the firm to the agents as a whole, thus ensuring an ex post constant surplus to the principal. (ii) Unlike before, the mechanism forces the two chosen agents to bear the principal's payment risk toward all other agents. (iii) All other agents' incentive compatibility and participation utility are preserved at the ex post level for all feasible reallocational manipulations by the coalition, including the two agents. Features (ii) and (iii) ensure the participation of noncollusive agents, which, along with (i), ensures the target level of surplus to the principal.[29] The last feature, (iii), requires ex post implementability of an outcome. Although that requirement limits the class of allocations/environments to which the above result applies, many plausible environments are known to be in that class. For instance, Mookherjee and Reichelstein (1992, Proposition 2) and Chung and Ely (2003, Proposition 4) provide the following sufficient condition.

LEMMA 3—Mookherjee–Reichelstein and Chung–Ely: *Suppose* $\Theta_i \equiv [\underline{\theta}_i, \overline{\theta}_i]$ (*i.e.*, *one-dimensional support*) *and* $\overline{U}_i(\theta_i) = \overline{U}_i$ *for all* $i \in N$. *Then, for any allo-*

[29]Given that the original mechanism is ex post implementable, a mechanism that satisfies (i) can be made ex post implementable for at most $n-2$ agents. This explains why at least two collusive agents need to be identified.

cation rule $q(\cdot)$ *such that* $\forall i, \forall \theta_i, \forall \theta_{-i}$,

(21) $$\frac{\partial}{\partial \theta_i} s_i(q(\theta_i', \theta_{-i}), \theta) \text{ is nonnegative and nondecreasing in } \theta_i'$$

there exists a transfer rule $t(\cdot)$ *such that* $M = (q, t)$ *is ex post implementable*.[30]

COROLLARY 3: *If any* $V \in \mathcal{V}$ *is implementable by* M *whose allocation rule satisfies the properties of Lemma 3, then, for any two agents* i *and* j, V *is implementable by a mechanism that is RCP with respect to any* C *that contains* i *and* j.

The assumed properties in Lemma 3 hold at the optimal mechanism in many well-known mechanism design problems, such as auctions, procurement, regulation, nonlinear pricing, and public goods provision.[31] Although the sufficient condition presumes a continuous type space, a discrete type problem can be reformulated as a continuous type problem without any loss (see Skreta (2006), for example). Hence, the result applies to all existing models of collusion.

REMARK 3—Correlated Types and Subgroup Collusion: We conjecture that a version of Theorem 3 holds generically in a large class of correlated types cases, with $|C| \geq 3$. Ex post implementability appears to be problematic, however. Even though ex post incentive compatibility alone seems feasible generically (see Mookherjee and Reichelstein (1992)), that requirement combined with ex post individual rationality is difficult to satisfy for the full-extraction first-best outcome. Whether and at what cost collusion by a subset of agents can be prevented remain an open question in the correlated type environment.

9. COLLUSION PROPOSED BY AN INFORMED AGENT

Previous sections have employed the LM modeling approach whereby a third party representative organizes a collusive agreement on behalf of members of the coalition. Even though this approach serves as a useful metaphor and we have allowed for a variety of scenarios within this approach, this modeling assumption may not be most descriptive of a typical collusion process. In a typical situation, a member of a coalition may initiate and propose a collusive

[30]Although Proposition 2 of Mookherjee and Reichelstein (1992), from which this lemma is adapted, does not prove ex post individual rationality, it is implied as Mookherjee and Reichelstein (1992) and Chung and Ely (2003) argue because the first condition implies that there is a single worst type.

[31]The connection with the literature can be made clearer with the following sufficient conditions due to Mookherjee and Reichelstein:

(i) One-dimensional condensation: There exist $h_i : \mathcal{Q} \to \mathbb{R}$ and $d_i(\cdot, \cdot) : \mathbb{R} \times \Theta \to \mathbb{R}$, twice differentiable, such that $s_i(q, \theta) = d_i(h_i(q), \theta)$.

(ii) Single crossing: $\partial^2 d_i / \partial h_i \partial \theta_i \geq 0$.

(iii) Ex post monotonicity: $\forall \theta_{-i} \in \Theta_{-i}$, $h_i(q(\theta_{-i}, \cdot))$ is nondecreasing.

agreement. We consider this latter scenario in the current model. The obvious difficulty with modeling this latter scenario is the "informed principal problem," since the agent who proposes a collusive scheme is privately informed of his type and may thus want to use a contract offer to signal his type to the other agent. The implications of such problems for coalition formation as well as for the principal's response to collusion have not been studied, except for the recent work by Quesada (2004). She finds the second-best outcome to be collusion-proof implementable in the LM setting with a binary type and a perfect complementarity technology.[32] Her result exploits special features of that setting,[33] however, leaving the generality of her result in question. We show below that our RCP design can be utilized to prevent collusion proposed by an informed agent in a much more general setting.[34]

To begin, we assume that there are only two agents and that agent 1 (instead of a third party representative) makes a take-it-or-leave-it collusion offer to the other agent at date $1\frac{1}{4}$. The rest of the structure remains the same as before. With this model, we show that there exists an RCP mechanism that implements the noncollusive optimal outcome in all equilibria supported by passive out-of-equilibrium beliefs (i.e., the beliefs invoked when agent 2 rejects agent 1's collusion offer). To this end, much like Quesada, we apply Maskin and Tirole's (1992, Theorem 1^*, p. 35) characterization of the informed principal problem, with agent 1 taking the role of the informed principal in their setup. Their characterization assumes two agents (i.e., one principal and one agent) and finite types distributed independently across the agents. Using their result thus requires us to restrict our model accordingly. Specifically, each agent $i = 1, 2$ draws a type independently from a finite set. We further assume that the second-best outcome is efficiently implementable via an ex post incentive compatible mechanism.[35] As noted in Lemma 3, this set includes most of the cases considered in the literature.

PROPOSITION 3: *Suppose there are only two agents*, *each with types drawn independently from a finite set*. *If there is a mechanism M^* that efficiently implements V^* and is ex post incentive compatible*, *then there exists an RCP mechanism*

[32]Quesada (2004) also considers ex ante collusion (on participation decisions) and finds that the second-best outcome is not collusion-proof implementable.

[33]The features of the LM model allow the second-best outcome to be collusion-proof implementable via ex post individually rational dominant strategies.

[34]A structurally similar problem is "resale" following an auction. Similarly to our problem, the winning bidder is informed about his type when he deals with the losing bidders in the resale market. In the resale problem, however, the (resale) contract proposal is made after the initial assignment, whereas the collusion proposal is made prior to the "play" of the grand mechanism. This difference turns out to matter. Since the bidders' types are updated after initial allocation, the crucial issue that faces the principal is how to "manipulate" the updating of types through initial assignment (see Zheng (2002)). This issue does not arise in our problem.

[35]This is weaker than assuming $V^* \in \mathcal{V}^{\mathrm{EP}}$, given that $\mathcal{V}^{\mathrm{EP}}$ also requires ex post individual rationality.

$\bar{M} \in \mathcal{M}$ that implements the noncollusive optimal payoff V^* for the principal in all perfect Bayesian equilibria supported by passive beliefs.

The proof is given in Appendix B.

Our focus on the equilibria supported by passive beliefs makes our approach comparable to weak collusion-proof implementation often invoked in the standard model of collusion. That restriction also makes our result nontrivial.[36] Indeed, our collusion-proof equilibrium exploits the RCP design of the mechanism and would not work without that feature. As before, our RCP mechanism makes the two agents residual claimants of the net social surplus—after paying off the principal V^*—achievable from any feasible collusive proposal that agent 1 may make. In addition, the mechanism satisfies ex post incentive compatibility for agent 2. It turns out that these two features guarantee the same payoffs for both agents in any equilibria supported by the passive beliefs, as in the noncollusive equilibrium. Hence, both agents participate in any such equilibria and the RCP feature then guarantees the optimal payoff V^* for the principal.[37]

Neither the argument of Proposition 3 nor the result of Maskin and Tirole appears to readily extend to the case of correlated types. Nevertheless, we offer another result that will be useful for that case. Consider our general model with $n \geq 2$ and an arbitrary type distribution. Given the condition of Lemma 2, the full-extraction first-best result is generically RCP implementable. The next result implies that such an outcome is collusion-proof implementable even when an informed agent proposes collusion.

PROPOSITION 4: *Suppose the principal offers a mechanism $M^* = (q^*, t^*) \in \mathcal{M}$ that satisfies (6) (and hence is RCP) and implements the first-best allocation, i.e.,*

$$q^*(\theta) \in \arg\max_{q \in \mathcal{Q}} v(q) + \sum_{i \in N} s_i(q, \theta) \quad \forall\, \theta \in \Theta. \tag{22}$$

Then there exists a perfect Bayesian equilibrium supported by passive beliefs in which agent 1 proposes the null side contract and outcome M^* is implemented.

[36]Without this restriction, it is not difficult to implement any feasible mechanism with collusion. For instance, as in Proposition 8* of Maskin and Tirole (1992), given any direct mechanism $M \in \mathcal{M}$, one can design an indirect mechanism that admits two equilibria: one where agents play M truthfully; the other where agent 2 just receives a very high payment from agent 1. One can then sustain a perfect Bayesian equilibrium where no collusive proposal is made, given the (out-of-equilibrium) belief that any rejection by agent 2 of such a proposal would trigger the second equilibrium to be much more unfavorable to agent 1. This latter belief is clearly implausible and makes the equilibrium less than believable.

[37]This argument does not use the fact that $n = 2$; neither does Theorem 1* of Maskin and Tirole (1992). We thus conjecture that the result holds for more than two agents.

See Appendix C for the proof.

This result also exploits the RCP design. As before, the mechanism used in the proof makes the agents residual claimants of net surplus achievable from any feasible collusive proposal. This feature makes any deviation from the efficient allocation rule unprofitable to the proposer, given passive beliefs.

10. HIERARCHICAL DELEGATION OF CONTRACTS

Part of an interest in studying collusion stems from the hope that it may offer to explain some organizational forms that are otherwise difficult to justify. Hierarchical delegation of contracts is one such organizational practice. If a principal delegates to, say agent 1, the authority to contract with agent 2, then the principal loses the opportunity to communicate with the latter and to choose his contract terms in her best interest. This control loss makes delegated contracting difficult to justify, despite its popularity. One can at best hope that delegated contracting does as well as centralized contracting—i.e., implements the second-best outcome. If agents' types are uncorrelated, Melumad, Mookherjee, and Reichelstein (1995) (hereafter, MMR) show that delegation achieves the second-best outcome if and only if, under delegation, (a) the principal monitors individual output contributions by all agents (q in our model) and (b) agent 1 can be compelled to make his participation decision before he communicates with agent 2.[38] Condition (a) is needed for the principal to be able to counteract a potential monopoly distortion that may arise from the increased bargaining power gained by agent 1. Condition (b) is needed for the individual rationality to hold at the interim level, so as to prevent agent 1 from commanding rents based on the information he learned about agent 2.

Does collusion make a difference? Although centralization still confers (weakly) more control to the principal than does delegation,[39] the latter seems relatively more attractive when the former is subject to collusion. To what extent collusion justifies hierarchical delegation has been the subject of much recent research, but no general answer has emerged yet. Some authors established equivalence in some cases (Laffont and Martimort (1998), Faure-Grimaud, Laffont, and Martimort (2003)) whereas others showed nonequivalence in other cases (Celik (2004), Mookherjee and Tsumagari (2004)). Our collusion-proof implementation results (under centralization) enable us to provide some general answer on the issue and fill an important gap in the literature.

[38]These conditions are "necessary" for equivalence in the sense that a counterexample can be found where failure of either condition leads to nonequivalence (see MMR). Mookherjee and Reichelstein (2001) also extend the sufficiency part to the case with any finite number of agents.

[39]Specifically, a centralized contract may enable the principal to manipulate agents' opportunity cost of participating in collusion since it will determine their status quo payoffs. The principal enjoys no such leverage relative to agent 2 under delegation.

Specifically, we have shown that the second-best outcome is achievable under centralization in the presence of collusion, whether it is organized by a third party (Corollary 1) or by an informed agent (Propositions 3 and 4). Hence, for delegation to do as well as centralization, the former must implement the second-best outcome. It then follows from MMR that, for uncorrelated types, delegation is inferior to centralization unless conditions (a) and (b) both hold. In other words, these conditions continue to be the relevant requirements for equivalence of the two arrangements, even in the presence of collusion.

This perspective explains some of the existing results. For instance, Laffont and Martimort's (1998) equivalence result follows immediately from MMR's conditions being satisfied in their model; in particular, their perfect complementarity technology makes it trivial for the principal to observe agents' individual outputs. Also consistent with our perspective, Mookherjee and Tsumagari (2004) showed delegation to be strictly inferior if agents can reallocate output and agent 1 can postpone his participation decision until after he communicates with agent 2, i.e., both conditions (a) and (b) fail.[40]

Thus far, no result is available for the case where only one of the MMR conditions holds. For instance, whether delegation is optimal is not known for the standard case where agents' participation constraints hold at the interim level but their technologies/preferences are general enough to admit nontrivial reallocational opportunities for the agents. Our results provide an unambiguous answer for this case. It follows from MMR that the agents' reallocational ability, when undetected by the principal, leads to a monopoly distortion under delegation. By contrast, the agents' reallocational ability does not prevent the principal from achieving the second-best outcome via a centralized contract. Hence, delegation is strictly inferior in this case.[41] In sum, our results suggest that hierarchical delegation is no more justifiable when the agents are collusive than when they are not, at least if their types are uncorrelated.

REMARK 4—Correlated Types and Delegation: Our result offers no general perspective when agents' types are correlated, since there is no MMR-like re-

[40]Our result does not imply theirs, however, since their model of collusion under centralization permits colluding agents to exit from the grand contract after communicating with each other, so their participation constraints hold ex post even for centralized contracting.

[41]The comparison could, in principle, depend on how one models collusion under centralized contracting. Laffont and Martimort (1998), for instance, invoke the third-party-initiated collusion, which treats the agents symmetrically in terms of their relative bargaining power. Since a delegated agent has the full bargaining power under delegation, the latter then involves a shift in bargaining power within the agents as well as the usual control loss for the principal. Mookherjee and Tsumagari (2004) adopt a different model where a collusive proposal is made by one agent (agent 1 in our discussion) in a take-it-or-leave-it fashion under centralization, so that the two formats differ only in terms of the principal's control loss. Regardless of the differences, our result implies that the second-best outcome is achievable under centralization. So, the nonequivalence result is quite robust.

sult for the no collusion benchmark. Faure-Grimaud, Laffont, and Martimort (2003) and Celik (2004) consider models in which a nonproductive supervisor observes an imperfect signal about the type of a productive agent. Differences in the informational structures have led them to reach different conclusions on the value of delegation. In both cases, however, conditions (a) and (b) are met. In particular, the nonproductive role of supervisor makes condition (a) trivial. If the supervisor also had a productive role, there could be additional distortion associated with delegation, rendering it unambiguously inferior.

11. CONCLUSION

We have shown that the optimal noncollusive mechanism can be made collusion-proof in a broad class of circumstances, including both uncorrelated and correlated types environments, in a way robust to the specifics of a coalition's objective, its manipulation technique, or its exact makeup. This result unifies several observations scattered in the literature and provides a general insight into how the transaction cost associated with agents' private information can be exploited to overcome collusion. An equally valuable lesson from the current paper may lie in furthering the understanding of the true scope of collusion. Although the mechanism we propose applies to a general class of technologies, preferences, and agents' type structures, it requires several important conditions. Recognizing these conditions can shed some light on the factors that can make collusion truly problematic.[42]

First, we followed the extensive form of LM (1997, 2000) in which a coalition is formed after the agents sign up for the principal's grand contract. This means that we do not allow the agents to collude on their participation decisions.[43] Although this assumption makes sense in many situations, there are circumstances in which agents may be able to collude prior to their participation decisions. To illustrate, consider our example in Section 2 and our collusion-proof mechanism that charges the agents 2/3. If they can collude prior to participating in such a mechanism, they may refuse to participate whenever their costs exceed 2/3, which will undermine the implementation of the second-best

[42]When these conditions fail, our method of collusion-proofing may not provide a useful guide for solving the collusion problem, and the traditional approach of optimizing within the class of collusion-proof mechanisms may again be useful. In this sense, the current paper complements the existing approach.

[43]This assumption may not be as restrictive as it may appear. Many forms of collusion that involve coordinated participation are replicable by a collusive arrangement in our model. For instance, McAfee and McMillan (1992) consider collusion that sends only one selected bidder to the official auction. This is replicated by an arrangement that sends all bidders, but all of them (except possibly for one) bid a reserve price. The good can be then reallocated to the selected winner, if necessary.

outcome. The extent and form of contract that can deal with such an early collusion remain an important question to study.[44]

Second, our collusion-proof implementation relies largely on the risk neutrality of the agents. An important feature of our mechanism is that it makes the agents residual claimants, which means shifting all payoff risks (i.e., the payoff variability) to the agents. Imposing such risks requires providing a risk premium to the agents if they are risk averse. Similarly, our mechanism may sometimes require positive entry fees, which agents may be either unwilling or unable to pay due to their risk aversion or liquidity constraints. Risk aversion and liquidity constraints will thus introduce a real trade-off in dealing with collusion.[45]

Third, our collusion-proof implementation relies on a Bayesian mechanism, which cannot generally be made either a dominant strategy or ex post implementable for all agents. As is often recognized, common knowledge required for Bayesian implementation is demanding. Relaxing this restriction will likely entail a real cost of preventing collusion. The exact nature of this cost and the method of minimizing it remain interesting open questions.

Dept. of Economics, Columbia University, 420 West 118th St., Room 1016 IAB, New York, NY 10027, U.S.A.; Dept. of Economics, University of Wisconsin-Madison, Madison, WI 53706, U.S.A.; yc2271@columbia.edu, www.columbia.edu/~yc2271.

and

Faculty of Economics, Yonsei University, Seoul 120-749, Korea; jikim@yonsei.ac.kr, eclass.yonsei.ac.kr/jikimdir.

Manuscript received May, 2004; final revision received February, 2006.

APPENDIX A: PROOF OF LEMMA 1

We use the following theorem.

THEOREM OF THE ALTERNATIVE—Fredholm[46]: *For a matrix A and a vector a, the linear system $Ax = a$ has a solution x^* if and only if, for any vector λ,* $\lambda A = \mathbf{0}$ implies $\lambda a = 0$.

[44]A few interesting papers have already employed an extensive form that permits agents to collude on their participation decisions. See Che and Kim (2005), Dequiedt (2004), Pavlov (2004), Quesada (2004), and Mookherjee and Tsumagari (2004).

[45]The logic is precisely the same as why "selling the firm to an agent" does not work in the traditional moral hazard model with a risk averse agent. Faure-Grimaud, Laffont, and Martimort indeed show the risk aversion can make dealing with collusion costly.

[46]See Carter (2001, p. 392), for instance.

Given this theorem, a solution to the system (12) exists if, for any (row) vectors $(\lambda_i^P, \lambda_i^B)_{i=1}^n$ and any (row) vector ξ,

$$\lambda_i^P P_i + \lambda_i^B B_i + \xi = \mathbf{0} \quad \forall\, i \quad \text{implies} \tag{23}$$
$$\sum_{i\in N} \lambda_i^P \cdot T_i + \sum_{i\in N} \lambda_i^B \cdot S_i + \xi \cdot [v - \Delta] = 0.$$

Note that λ_i^P, λ_i^B, and ξ are of sizes ℓ_i, $\ell_i(\ell_i - 1)$, and L, respectively.

To prove (23), suppose $\lambda_i^P P_i + \lambda_i^B B_i + \xi = \mathbf{0}$ for all i, which implies

$$\lambda_i^P P_i + \lambda_i^B B_i = \lambda_j^P P_j + \lambda_j^B B_j = -\xi. \tag{24}$$

Let agents i and j be the ones that satisfy Condition (PI′). The condition means that the space spanned by row vectors of P_i and B_i has only one-dimensional vector space in common with the space spanned by row vectors of P_j and B_j. According to (24), ξ must belong to this one-dimensional space. However, we have

$$e_1' P_1 = \cdots = e_n' P_n = \mu^{0'},$$

where e_i is the (column) vector of size ℓ_i whose elements are all 1's. Thus, it must be that for some scalar β, $\xi = \beta \mu^{0'}$. Then

$$\begin{aligned}
\sum_{i\in N} \lambda_i^P \cdot T_i &= \sum_{i\in N} \lambda_i^P P_i \cdot t_i = -\sum_{i\in N} [\lambda_i^B B_i + \xi] \cdot t_i \\
&= -\sum_{i\in N} \lambda_i^B \cdot S_i - \beta \sum_{i\in N} \mu^{0'} \cdot t_i \\
&= -\sum_{i\in N} \lambda_i^B \cdot S_i - \beta \sum_{i\in N} \mathbb{E}[t_i(\tilde{\theta})].
\end{aligned}$$

Also,

$$\begin{aligned}
\xi \cdot [v - \Delta] &= \beta \mu^{0'} \cdot [v - \Delta] \\
&= \beta \mu^{0'} \cdot v - \beta \mathbb{E}\big[v(q(\tilde{\theta}))\big] + \beta \sum_{i\in N} \mathbb{E}[t_i(\tilde{\theta})] \\
&= \beta \sum_{i\in N} \mathbb{E}[t_i(\tilde{\theta})].
\end{aligned}$$

Therefore, we have

$$\sum_{i\in N} \lambda_i^P \cdot T_i + \sum_{i\in N} \lambda_i^B \cdot S_i + \xi \cdot [v - \Delta]$$

$$= -\sum_{i\in N}\lambda_i^B\cdot S_i - \beta\sum_{i\in N}\mathbb{E}[t_i(\tilde\theta)] + \sum_{i\in N}\lambda_i^B\cdot S_i + \beta\sum_i\mathbb{E}[t_i(\tilde\theta)]$$
$$= 0,$$

proving (23). *Q.E.D.*

APPENDIX B: PROOF OF PROPOSITION 3

As stated in the text, our proof applies Theorem 1* of Maskin and Tirole (1992). This requires developing two welfare concepts. To this end, we define several notations. For each $\theta_i \in \Theta_i$, let $p_i^0(\theta_i)$ denote the probability that agent $i = 1, 2$ realizes that type. As before, $\mu := (p_1(\cdot), p_2(\cdot))$ denotes an arbitrary prior distribution of types and $\mu_i := p_i(\cdot)$ denotes the prior for agent $i = 1, 2$. We reserve $\mu^0 := (p_1^0(\cdot), p_2^0(\cdot))$ and $\mu_i^0 := p_i^0(\cdot)$ for true priors. Let

$$u_i^M(\tilde\theta_1, \tilde\theta_2|\theta_1, \theta_2) := s_i(q(\tilde\theta_1, \tilde\theta_2), \theta_1, \theta_2) + t(\tilde\theta_1, \tilde\theta_2)$$

denote agent i's ex post payoff from mechanism $M = (q, t)$ when the agents have types (θ_1, θ_2) but report $(\tilde\theta_1, \tilde\theta_2)$. For each $\theta_1 \in \Theta_1$, we let $M_{\theta_1} := (q(\theta_1, \cdot), t(\theta_1, \cdot))$ denote a component of a menu that corresponds to a report of type θ_1 by agent 1. Hence, we can write $M = \{M_{\theta_1}\}_{\theta_1\in\Theta_1}$.

As before, we fix an arbitrary grand mechanism $M = (q, t)$ offered by the principal and consider a reallocational manipulation of M proposed by agent 1. We first define so-called *interim efficiency*. A reallocational manipulation of M, $\check M$, is said to be *interim efficient* (IE*) *relative to prior* $\hat\mu_1$ if, for some $(w(\theta_1))_{\theta_1\in\Theta_1} \in \mathbb{R}_{++}^{|\Theta_1|}$, $\check M$ solves

(IE*$(\hat\mu_1; M)$)

$$\max_{\tilde M\in\mathcal{RM}_M}\sum_{\theta_1\in\Theta_1} w(\theta_1)\Bigg(\sum_{\theta_2\in\Theta_2} p_2^0(\theta_2)u_1^{\tilde M}(\theta_1, \theta_2|\theta_1, \theta_2)\Bigg),$$

(IC^1)
$$\sum_{\theta_2\in\Theta_2} p_2^0(\theta_2)u_1^{\tilde M}(\theta_1, \theta_2|\theta_1, \theta_2) \geq \sum_{\theta_2\in\Theta_2} p_2^0(\theta_2)u_1^{\tilde M}(\theta_1', \theta_2|\theta_1, \theta_2) \quad \forall\,\theta_1, \theta_1' \in \Theta_1,$$

(IC$^2(\hat\mu_1)$)
$$\sum_{\theta_1\in\Theta_1} \hat p_1(\theta_1)u_2^{\tilde M}(\theta_1, \theta_2|\theta_1, \theta_2) \geq \sum_{\theta_1\in\Theta_1} \hat p_1(\theta_1)u_2^{\tilde M}(\theta_1, \theta_2'|\theta_1, \theta_2) \quad \forall\,\theta_2, \theta_2' \in \Theta_2,$$

($\mathrm{IR}^2_{\check{M}}(\hat{\mu}_1)$)

$$\sum_{\theta_1\in\Theta_1}\hat{p}_1(\theta_1)u_2^{\tilde{M}}(\theta_1,\theta_2|\theta_1,\theta_2) \geq \sum_{\theta_1\in\Theta_1}\hat{p}_1(\theta_1)u_2^{\check{M}}(\theta_1,\theta_2|\theta_1,\theta_2) \quad \forall\,\theta_2\in\Theta_2.$$

Next, a reallocational manipulation of M, $\tilde{M}^{\mathrm{RSW}}(M)$, is said to be *RSW* relative to M* if $\tilde{M}^{\mathrm{RSW}}(M)=\bar{M}$ and, for each $\theta_1\in\Theta_1$, there exists a mechanism $\{\bar{M}_{\theta_1},\check{M}_{\theta_1'}\}_{\theta_1'\in\Theta_1\setminus\{\theta_1\}}$ that solves

($\mathrm{RSW}^*_{\theta_1}(M)$)

$$\max_{\tilde{M}\in\mathcal{RM}_M}\sum_{\theta_2\in\Theta_2}p_2^0(\theta_2)u_1^{\tilde{M}}(\theta_1,\theta_2|\theta_1,\theta_2)$$

subject to (IC^1),

(EPIC^2)

$$u_2^{\tilde{M}}(\tilde{\theta}_1,\theta_2|\tilde{\theta}_1,\theta_2)\geq u_2^{\tilde{M}}(\tilde{\theta}_1,\theta_2'|\tilde{\theta}_1,\theta_2) \quad \forall\,\tilde{\theta}_1\in\Theta_1, \quad \forall\,\theta_2,\theta_2'\in\Theta_2,$$

(EPIR^2_M)

$$u_2^{\tilde{M}}(\tilde{\theta}_1,\theta_2|\tilde{\theta}_1,\theta_2)\geq u_2^{M}(\tilde{\theta}_1,\theta_2|\tilde{\theta}_1,\theta_2) \quad \forall\,\tilde{\theta}_1\in\Theta_1, \quad \forall\theta_2\in\Theta_2.$$

Theorem 1* of Maskin and Tirole (1992) proves that, if a RSW* mechanism is IE* relative to some positive prior, then any mechanism that satisfies (IC) and (IR), and weakly Pareto-dominates the RSW* for agent 1 is supported as an equilibrium of the game where agent 1 proposes a contract to agent 2. We apply this result to prove Proposition 3. By the hypothesis, there exists an ex post incentive compatible mechanism $M^*=(q^*,t^*)$ that implements V^* efficiently. We now construct mechanism $\bar{M}=(q^*,\bar{t})$, where

$$\bar{t}_2(\theta):=t_2^*(\theta)+\rho(\theta_1)-\mathbb{E}_{\tilde{\theta}_1}[\rho(\tilde{\theta}_1)], \tag{25}$$

where

$$\rho(\theta_1):=\mathbb{E}_{\tilde{\theta}_2}\left[v(q^*(\theta_1,\tilde{\theta}_2))-\sum_{i=1,2}t_i^*(\theta_1,\tilde{\theta}_2)\right]$$

and, for agent 1,

$$\bar{t}_1(\theta):=-\bar{t}_2(\theta)+v(q^*(\theta))-\mathbb{E}_{\tilde{\theta}_1}[\rho(\tilde{\theta}_1)]. \tag{26}$$

As can be seen from (26), the transfers sum to a level that ensures an ex post constant payoff of $\mathbb{E}_{\tilde{\theta}_1}[\rho(\tilde{\theta}_1)] = V^*$ to the principal. Further, these transfers ensure the same interim payoffs for both agents and the same ex post incentive for agent 2, as t^*. Hence, $\bar{M}$ is RCP.

Suppose the principal offers $\bar{M}$. By construction, $\bar{M}$ is ex post incentive compatible for agent 2. This means that $\bar{M}$ satisfies the constraints of ($\mathrm{RSW}^*_{\theta_1}(\bar{M})$) for each $\theta_1 \in \Theta_1$. Hence, each type θ_1 of agent 1 can guarantee an interim payoff of $U_i^{\bar{M}}(\theta_1)$ by offering $\bar{M}$, i.e., the null side contract, implying that the payoff for each type of agent 1 from an RSW* mechanism relative to $\bar{M}$ must be at least that of $\bar{M}$. At the same time, the RSW* mechanism relative to $\bar{M}$ must be a reallocational manipulation of $\bar{M}$, so it gives V^* to the principal (by design of $\bar{t}$), and it must satisfy (IC) and (IR),[47] so it must be noncollusive optimal. Since $\bar{M}$ has the same allocation rule as M^*, $\bar{M}$ must also implement V^* efficiently. Then the RSW* payoff for each type of agent 1 must equal that of $\bar{M}$. Otherwise, there must be a reallocational manipulation of $\bar{M}$ that gives strictly higher payoff to some type of agent 1 and no lower payoff to all other types of agent 1 and all types of agent 2 (since it must satisfy ($\mathrm{EPIR}^2_{\bar{M}}$)) than $\bar{M}$ does. Since $\bar{M}$ is RCP, this implies there exists a mechanism that implements V^* but generates higher total surplus than $\bar{M}$, which contradicts the fact that $\bar{M}$ efficiently implements V^*. We thus conclude that any RSW* allocation relative to $\bar{M}$ must yield the same payoff as $\bar{M}$ for each type of agent 1.

Next, we prove that $\bar{M}$ is interim efficient relative to some positive prior. To this end, consider another program,

$$(\mathrm{IE}^0) \qquad \max_{\tilde{M}\in\mathcal{RM}_{\bar{M}}} \sum_{\theta_1\in\Theta_1} p_1^0(\theta_1)\bigg(\sum_{\theta_2\in\Theta_2} p_2^0(\theta_2)u_1^{\tilde{M}}(\theta_1,\theta_2|\theta_1,\theta_2)\bigg),$$

subject to (IC^1), ($\mathrm{IC}^2(\mu_1^0)$), ($\mathrm{IR}^2_{\bar{M}}(\mu_1^0)$), and

$$(\mathrm{IR}^1) \qquad \sum_{\theta_2\in\Theta_2} u_1^{\tilde{M}}(\theta_1,\theta_2|\theta_1,\theta_2)p_2^0(\theta_2) \geq \overline{U}_1(\theta_1) \quad \forall\, \theta_1\in\Theta_1.$$

We claim that $\bar{M}$ solves (IE^0). First, since $\bar{M}$ is noncollusive optimal and RCP, any reallocational manipulation of $\bar{M}$ guarantees V^* to the principal. If there exists a mechanism $\tilde{M} \in \mathcal{RM}_{\bar{M}}$ that solves (IE^0) and yields agent 1 a higher (ex ante) payoff than $\bar{M}$, then $\tilde{M}$ is feasible and Pareto-dominates $\bar{M}$,

[47]That the RSW* mechanism satisfies (IC^1) can be checked easily and is established in Proposition 1 of Maskin and Tirole (1992). That it satisfies (IC^2) follows from (EPIC^2), which is required for ($\mathrm{RSW}^*_{\theta_1}(\bar{M})$). For each $\theta_1 \in \Theta_1$, a solution to ($\mathrm{RSW}^*_{\theta_1}(\bar{M})$) must satisfy (IR) since it satisfies ($\mathrm{EPIR}^2_{\bar{M}}$), and we conclude that it must give at least the payoff of $U_1^{\bar{M}}(\theta_1) \geq \overline{U}_1(\theta_1)$ to agent 1 with type $\theta_1 \in \Theta_1$.

since the former yields agent 2 no less payoff than the latter (due to a constraint in (IE^0)) and yields the principal V^* (since $\bar{M}$ is RCP), which contradicts that $\bar{M}$ efficiently implements V^*. Thus, $\bar{M}$ must solve (IE^0). Let $\lambda(\theta_1) \geq 0$ denote the Lagrangian multiplier associated with the constraint (IR^1) for type $\theta_1 \in \Theta_1$ at the solution. Then (IE^0) can be rewritten as

$$\max_{\tilde{M} \in \mathcal{RM}_{\bar{M}}} \sum_{\theta_1 \in \Theta_1} (p_1^0(\theta_1) + \lambda(\theta_1)) \left(\sum_{\theta_2 \in \Theta_2} p_2^0(\theta_2) u_1^{\tilde{M}}(\theta_1, \theta_2 | \theta_1, \theta_2) \right)$$

subject to

$$(\mathrm{IC}^1), (\mathrm{IC}^2(\mu_1^0)), \text{ and } (\mathrm{IR}_{\bar{M}}^2(\mu_1^0)),$$

from which it follows that $\bar{M}$ solves (IE$^*(\mu_1^0; \bar{M})$) for $w_1(\theta_1) := p_1^0(\theta_1) + \lambda(\theta_1)$, $\theta_1 \in \Theta_1$. We thus conclude that $\bar{M}$ is IE* relative to the true prior μ_1^0, which is positive.

Given the grand mechanism $\bar{M}$, let $\bar{\mathcal{M}}$ denote the set of all equilibrium outcomes supported by the passive belief. Theorem 1* of Maskin and Tirole (1992) states that any outcome $M \in \bar{\mathcal{M}}$ satisfies (IC^1), (IC$^2(\mu_1^0)$), and (IR$_{\bar{M}}^2(\mu_1^0)$), and also weakly Pareto-dominates RSW* allocation, that is, $U_1^M(\theta_1) \geq U_1^{\bar{M}}(\theta_1)$ $\forall\, \theta_1 \in \Theta_1$. Hence, any such equilibrium outcome induces both agents to participate in the grand mechanism $\bar{M}$, which yields the payoff of V^* for the principal, given the RCP feature of $\bar{M}$. *Q.E.D.*

APPENDIX C: PROOF OF PROPOSITION 4

The equilibrium strategies, given grand mechanism M^*, are described as follows: Agent 1 proposes the null side contract and the contract is accepted by all other agents, after which each agent reports truthfully in M^*. If agent 1 offers a nonnull contract, then each agent best responds to the following off-the-equilibrium belief: (1) Following agent 1's deviation, each agent $i \neq 1$ believes that agent 1 is of such a type that would benefit strictly from deviation and that any other agent, agent j $(\neq 1, i)$, will accept the deviation offer if he is of such a type that would be (weakly) better off from accepting it. (2) Following a rejection of any side contract proposed by agent 1, each agent holds the passive belief and reports truthfully when playing M^*, so M^* is truthfully implemented.

We show that there is no profitable deviation for agent 1. Suppose to the contrary that a positive measure of types of agent 1 is better off deviating to propose a nonnull side contract. For the deviation to be profitable, it must be accepted by a positive measure of other agents' types that get weakly better off by doing so, given the specified belief following that deviation. Let $\overline{\Theta}_1$ denote the set of agent 1's types that are deviating and let $\overline{\Theta}_i$ with $i \neq 1$ denote the set of agent i's types that are accepting the deviation. Let $\overline{\Theta} := \times_{i=1}^n \overline{\Theta}_i$, and define $\overline{\Theta}_{-i}$ and $\overline{\Theta}_{-i-j}$ as usual. Also, let $\tilde{M} = (\tilde{q}, \tilde{t}) \in \mathcal{RM}_{M^*}$ denote the mechanism/outcome being implemented via the deviation side contract. Let $\bar{u}_i^M(\theta) := s_i(q(\theta), \theta) + t_i(\theta)$ denote the ex post payoff that arises from outcome $M = (q, t)$.

For the deviation to be profitable for agent 1 with type $\theta_1 \in \overline{\Theta}_1$, we must have

$$\mathbb{E}\big[\bar{u}_1^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-1}\in\overline{\Theta}_{-1}\}} + \bar{u}_1^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-1}\notin\overline{\Theta}_{-1}\}}\big|\tilde{\theta}_1 = \theta_1\big] > \mathbb{E}[\bar{u}_1^{M^*}(\tilde{\theta})|\tilde{\theta}_1 = \theta_1]. \tag{27}$$

To understand the second term of the left-hand side, note that the deviation is rejected if $\tilde{\theta}_{-1} \notin \overline{\Theta}_{-1}$ and, given the belief in (2), M^* is truthfully implemented whenever a rejection occurs. Rewrite (27) as

$$\mathbb{E}\big[\bar{u}_1^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-1}\in\overline{\Theta}_{-1}\}}\big|\tilde{\theta}_1 = \theta_1\big] > \mathbb{E}\big[\bar{u}_1^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-1}\in\overline{\Theta}_{-1}\}}\big|\tilde{\theta}_1 = \theta_1\big].$$

Taking expectations across all types in $\overline{\Theta}_1$, we obtain

$$\mathbb{E}\big[\bar{u}_1^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\big] > \mathbb{E}\big[\bar{u}_1^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\big]. \tag{28}$$

For agent $i \neq 1$ with type $\theta_i \in \overline{\Theta}_i$ to accept the side contract, we must have[48]

$$\begin{aligned}&\mathbb{E}\big[\bar{u}_i^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-i}\in\overline{\Theta}_{-i}\}} + \bar{u}_i^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_1\in\overline{\Theta}_1, \tilde{\theta}_{-1-i}\notin\overline{\Theta}_{-1-i}\}}\big|\tilde{\theta}_i = \theta_i\big]\\ &\quad\geq \mathbb{E}\big[\bar{u}_i^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_1\in\overline{\Theta}_1\}}\big|\tilde{\theta}_i = \theta_i\big],\end{aligned}$$

which can be rewritten as

$$\mathbb{E}\big[\bar{u}_i^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-i}\in\overline{\Theta}_{-i}\}}\big|\tilde{\theta}_i = \theta_i\big] \geq \mathbb{E}\big[\bar{u}_i^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}_{-i}\in\overline{\Theta}_{-i}\}}\big|\tilde{\theta}_i = \theta_i\big].$$

Taking expectations across all types in $\overline{\Theta}_i$ yields

$$\mathbb{E}\big[\bar{u}_i^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\big] \geq \mathbb{E}\big[\bar{u}_i^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\big]. \tag{29}$$

Summing (28) and (29) across all agents, we obtain

$$\mathbb{E}\bigg[\sum_{i\in N}\bar{u}_i^{\tilde{M}}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\bigg] > \mathbb{E}\bigg[\sum_{i\in N}\bar{u}_i^{M^*}(\tilde{\theta})\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\bigg],$$

[48]This inequality can be explained in a similar way to (27). Here, the right-hand side and the second term of the left-hand side follow from the fact that agent i's belief in (1) is correct about what types of agent 1 would make the deviation offer and what types of each agent $j \neq 1, i$ would accept or reject it.

which implies that

$$\mathbb{E}\left[\left(v(\tilde{q}(\tilde{\theta})) + \sum_{i\in N} s_i(\tilde{q}(\tilde{\theta}), \tilde{\theta})\right)\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\right]$$
$$> \mathbb{E}\left[\left(v(q^*(\tilde{\theta})) + \sum_{i\in N} s_i(q^*(\tilde{\theta}), \tilde{\theta})\right)\mathbb{1}_{\{\tilde{\theta}\in\overline{\Theta}\}}\right],$$

since M^* satisfies (6) (or (11)) and $\tilde{M} \in \mathcal{RM}_{M^*}$. This contradicts (22). *Q.E.D.*

REFERENCES

ABDULKADIROĞLU, A., AND K.-S. CHUNG (2003): "Auction Design with Tacit Collusion," Unpublished Manuscript, Columbia and Northwestern Universities.[1063]

AOYAGI, M. (2003): "Bid Rotation and Collusion in Repeated Auctions," *Journal of Economic Theory*, 112, 79–105. [1063]

ARROW, K. (1979): "The Property Rights Doctrine and Demand Revelation under Incomplete Information," in *Economics and Human Welfare*, ed. by M. Boskin. New York, NY: Academic Press, 23–39.[1081]

BALDER, E. J. (1996): "On the Existence of Optimal Contract Mechanisms for Incomplete Information Principal–Agent Models," *Journal of Economic Theory*, 68, 133–148.[1071]

BALIGA, S., AND T. SJÖSTRÖM (1998): "Decentralization and Collusion," *Journal of Economic Theory*, 83, 196–232.[1063]

BARON, D. P., AND R. B. MYERSON (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930. [1068]

BOSE, S., E. OZDENOREN, AND A. PAPE (2005): "Optimal Auctions with Ambiguity," Unpublished Manuscript, University of Michigan.[1081]

BLUME, A., AND P. HEIDHUES (2002): "Modeling Tacit Collusion in Auctions," Unpublished Manuscript, University of Pittsburgh.[1063]

BRUSCO, S., AND G. LOPOMO (2002): "Collusion via Signalling in Simultaneous Ascending Bid Auctions with Heterogeneous Objects, with and Without Complementarities," *Review of Economic Studies*, 69, 407–436. [1063]

CAILLAUD, B., AND P. JEHIEL (1998): "Collusion in Auctions with Externalities," *RAND Journal of Economics*, 29, 680–702.[1063]

CARTER, M. (2001): *Foundations of Mathematical Economics*. Cambridge, MA: MIT Press. [1098]

CELIK, G. (2004): "Mechanism Design with Collusive Supervision," Unpublished Manuscript, University of British Columbia.[1063,1066,1095,1097]

CHE, Y.-K., AND J. KIM (2004): "Collusion-Proof Implementation of Optimal Mechanisms," SSRI Working Paper 2004-04, University of Wisconsin.[1068]

——— (2005): "Optimal Collusion-Proof Auctions," Unpublished Manuscript, Columbia University.[1098]

CHUNG, K.-S., AND J. ELY (2003): "Ex-post Incentive Compatible Mechanism Design," Unpublished Manuscript, Northwestern University.[1091,1092]

CRÉMER, J. (1996): "Manipulations by Coalitions under Asymmetric Information: The Case of Groves Mechanisms," *Games and Economic Behavior*, 13, 39–73. [1063]

CRÉMER, J., AND R. P. MCLEAN (1985): "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands Are Interdependent," *Econometrica*, 53, 345–362. [1081]

——— (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257. [1081,1086]

D'ASPREMONT, C., AND L. A. GÉRARD-VARET (1979): "Incentives and Incomplete Information," *Journal of Public Economics*, 11, 25–45.[1081]

DEQUIEDT, V. (2004): "Optimal Collusion and Optimal Auctions," Unpublished Manuscript, University of Toulouse.[1076,1098]

ESÖ, P., AND G. FUTOS (1999): "Auction Design with a Risk Averse Seller," *Economics Letters*, 65, 71–74.[1081]

ESÖ, P., AND J. SCHUMMER (2004): "Bribing and Signalling in Second Price Auctions," *Games and Economics Behavior*, 47, 299–324. [1063]

FAURE-GRIMAUD, A., J.-J. LAFFONT, AND D. MARTIMORT (2003): "Collusion, Delegation and Supervision with Soft Information," *Review of Economic Studies*, 70, 253–279.[1063,1066,1095, 1097]

FUDENBERG, D., D. K. LEVINE, AND E. MASKIN (1994): "The Folk Theorem with Imperfect Information," *Econometrica*, 62, 997–1039. [1084]

——— (1995): "Balanced-Budget Mechanisms with Incomplete Information," Unpublished Manuscript, UCLA.[1084]

GRAHAM, D. A., AND R. C. MARSHALL (1987): "Collusive Bidder Behavior at Single-Object Second-Price and English Auctions," *Journal of Political Economy*, 95, 1217–1239.[1063]

GREEN, J., AND J.-J. LAFFONT (1979): "On Coalition Incentive Compatibility," *Review of Economic Studies*, 44, 243–254. [1063]

JEHIEL, P., AND B. MOLDOVANU (2001): "Efficient Design with Interdependent Valuations," *Econometrica*, 69, 1237–1259. [1079]

JEHIEL, P., B. MOLDOVANU, AND E. STACCHETTI (1999): "Multimensional Mechanism Design for Auctions with Externalities," *Journal of Economic Theory*, 85, 258–293. [1069]

JEON, D. S., AND D. MENICUCCI (2005): "Optimal Second-Degree Price Discrimination and Arbitrage: On the Role of Asymmetric Information Among Buyers," *RAND Journal of Economics*, 36, 337–360.[1064,1073,1076]

KOSENOK, G., AND S. SEVERINOV (2004): "Individually Rational, Balanced-Budget Bayesian Mechanisms and the Informed Principal Problem," Unpublished Manuscript, Duke University.[1084,1085]

LAFFONT, J.-J., AND D. MARTIMORT (1997): "Collusion under Asymmetric Information," *Econometrica*, 65, 875–911. [1063,1064,1067,1097]

——— (1998): "Collusion and Delegation," *Rand Journal of Economics*, 29, 280–305.[1066,1095, 1096]

——— (2000): "Mechanism Design with Collusion and Correlation," *Econometrica*, 68, 309–342.[1063-1065,1067,1069,1081,1085,1097]

MAILATH, G., AND P. ZEMSKY (1991): "Collusion in Second Price Auctions with Heterogeneous Bidders," *Games and Economic Behavior*, 3, 467–486. [1063]

MARSHALL, R. C., AND L. M. MARX (2004): "Bidder Collusion," Unpublished Manuscript, Penn State University.[1063]

MASKIN, E., AND J. TIROLE (1992): "The Principal–Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, 60, 1–42.[1093,1094,1100-1103]

MCAFEE, R. P., AND J. MCMILLAN (1992): "Bidding Rings," *American Economic Review*, 82, 579–799.[1063,1067,1097]

MELUMAD, N. D., D. MOOKHERJEE, AND S. REICHELSTEIN (1995): "Hierarchical Decentralization of Incentive Contracts," *RAND Journal of Economics*, 26, 654–672.[1066,1095]

MOOKHERJEE, D., AND S. REICHELSTEIN (1992): "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules," *Journal of Economic Theory*, 56, 378–399. [1091,1092]

——— (2001): "Incentives and Coordination in Hierarchies," *Advances in Theoretical Economics*, 1, article 4. [1095]

MOOKHERJEE, D., AND M. TSUMAGARI (2004): "The Organization of Supplier Networks: Effects of Delegation and Intermediation," *Econometrica*, 72, 1179–1219. [1063,1066,1073,1074,1076, 1095,1096,1098]

MYERSON, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73. [1066]

MYERSON, R. B., AND M. SATTERTHWAITE (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 59, 451–465. [1065]

PAVLOV, G. (2004): "Colluding on Participation Decisions," Unpublished Manuscript, Boston University.[1098]

QUESADA, L. (2004): "Modeling Collusion as an Informed Principal Problem," Unpublished Manuscript, University of Wisconsin.[1064,1073,1093,1098]

SEVERINOV, S. (2003): "Optimal Organization: Centralization, Decentralization or Delegation?" Unpublished Manuscript, Duke University.[1063]

SKRETA, V. (2006): "Mechanism Design for Arbitrary Type Spaces," *Economics Letters*, 91, 293–299.[1092]

SKRZYPACZ, A., AND H. HOPENHAYN (2004): "Tacit Collusion in Repeated Auctions," *Journal of Economic Theory*, 114, 153–169. [1063]

TIROLE, J. (1986): "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organization*, 2, 181–214.[1063]

ZHENG, C. (2002): "Optimal Auction with Resale," *Econometrica*, 70, 2197–2224. [1093]