Mediated Partnerships*

David Rahman† and Ichiro Obara
University of Minnesota and UCLA

First Submission: November 6, 2005.
This Draft: April 6, 2009.

Abstract

This paper studies partnerships that employ a mediator to improve their contractual ability. Intuitively, profitable deviations must be attributable, i.e., there must be some group behavior such that an individual can be statistically identified as innocent, to provide incentives in partnerships. Mediated partnerships add value by effectively using different behavior to attribute different deviations. As a result, mediated partnerships are necessary to provide the right incentives in a wide range of economic environments.

JEL Classification: D21, D23, D82.
Keywords: mediated contracts, partnerships, private monitoring.

*Many thanks are owed to Harold Demsetz, Larry Jones, Michihiro Kandori, Narayana Kocherlakota, David Levine, Roger Myerson, Itai Sher, Joe Ostroy, Phil Reny, Joel Sobel, Bill Zame, co-editor Larry Samuelson and three anonymous referees for help with previous drafts. We are also grateful to numerous seminar audiences.

†Financial support from the Spanish Ministry of Education's Research Grant No. SEJ 2004-07861 while at Universidad Carlos III de Madrid is gratefully acknowledged.

1 Introduction

Providing incentives in partnerships is a classic topic of economic theory.[1] Although it is well-known that communication is a basic facet of incentive provision (Aumann, 1974; Myerson, 1986), this insight has not been systematically applied to partnership problems. This paper adds to the literature by asking the following question. Consider a group of individuals whose behavior is subject to moral hazard, but with rich communication and contractual protocols: access to (i) a disinterested mediator that can make confidential, verifiable but non-binding recommendations to agents, and (ii) budget-balanced payment schemes[2] that may depend on both the mediator's recommendations and individual reports. What outcomes can this group enforce?

Our main result (Theorem 1) shows that identifying obedient agents (IOA) is both necessary and sufficient for every outcome to be virtually enforceable[3] in this mediated environment, regardless of preferences. IOA means that for any profile of deviations, there is some behavior by the agents that statistically identifies an innocent individual after any unilateral deviation in the profile. IOA enjoys the following crucial property: different behavior may be used to attribute innocence after different deviations.

Let us intuitively explain this result. On the one hand, providing incentives with budget balance requires punishing some agents and rewarding others simultaneously. If after a unilateral deviation an innocent party cannot be identified then the deviator could have been anyone, so the only way to discourage the deviation is to punish everyone. However, this violates budget balance. On the other hand, IOA implies that budget-balanced incentives can be provided by rewarding the innocent and punishing all others. Here we make use of the mediator as follows. By the Minimax Theorem, there is a contract that discourages every deviation profile simultaneously if and only if for every deviation profile there is a contract that discourages it, where now different contracts may be employed to discourage different deviation profiles. Rich contractual protocols enable the use of payments that differ after different recommended actions, so in order to reward the innocent after a given deviation profile, different behavior may be used to find such innocent parties. But this is just the definition of IOA.

[1]See Alchian and Demsetz (1972), Holmström (1982), Radner et al. (1986), Legros and Matsushima (1991), Legros and Matthews (1993), d'Aspremont and Gérard-Varet (1998) and others.

[2]Budget balance means that the sum of payments across individuals always equals zero.

[3]An outcome is "virtually enforceable" if there is an enforceable outcome arbitrarily close to it.

The value of mediated partnerships over ordinary ones (Theorems 2 and 4) now follows. Without payment schemes contingent on recommendations, it is possible to provide incentives by rewarding the innocent only if the same behavior is used to attribute innocence after every deviation. The difference between this requirement and the clearly less stringent IOA characterizes the value of mediated partnerships. As it turns out, mediated partnerships provide incentives in many natural environments where incentives would otherwise fail. For instance, for generic distributions of output, mediated partnerships can provide incentives (Theorem 8) even without production complementarities, yet ordinary ones cannot (Example 1).[4]

This paper adds to the literature (Section 6) in two basic ways. Firstly, it extends the work of Legros and Matthews (1993), who derived nearly efficient partnerships in restricted environments with output-contingent contracts. Although they noted that identifying the innocent is important for budget-balanced incentives, they did not address statistical identification, and did not use different behavior to identify the innocent after different deviations. Secondly, being necessary for Theorem 1, IOA exhausts the informational economies from identifying the innocent rather than the guilty.[5] This contrasts the literature on repeated games, where restricted communication protocols were used by Kandori (2003) and others to prove the Folk Theorem.[6] Such papers typically require a version of pairwise full rank (Fudenberg et al., 1994), which intuitively means identifying the deviator after every deviation. This is clearly more restrictive than IOA, which only requires identifying a non-deviator.

The paper is organized as follows. Section 2 presents a motivating example where a mediated partnership is virtually enforced, yet none of the papers above apply. Section 3 presents the model and main definitions. Section 4 states our main results, discussed above. Section 5 makes three comments: (i) it refines our main assumptions in the specific context of public monitoring, (ii) establishes genericity of IOA (Theorem 8), and (iii) includes participation and liability constraints in the model. Section 6 reviews the literature on contract theory and repeated games and compares it to this paper. Finally, Section 7 concludes. Proofs appear in Appendix A.

[4]For example, we do not require that the distribution of output has a "moving support," i.e., the support of the distribution depends on individual behavior. This assumption, made by Legros and Matthews (1993), is not generic, so an arbitrarily small change in probabilities leads to its failure.

[5]Heuristically, knowing who deviated implies knowing someone who did not deviate, but knowing someone who did not deviate does not necessarily imply knowing who did.

[6]See Section 6 for a more detailed discussion of this literature.

2 Example

We begin our analysis of mediated partnerships with an example to capture the intuition behind our main result, Theorem 1. The example suggests the following intuitive way of attaining a "nearly efficient" partnership: appoint a *secret principal*.

Example 1. Consider a fixed group of n individuals. Each agent i can either work ($a_i = 1$) or shirk ($a_i = 0$). Let $c > 0$ be each individual's cost of effort. Effort is not observable. Output is publicly verifiable and can be either good (g) or bad (b). The probability of g equals $P(\sum_i a_i)$, where P is a strictly increasing function of the sum of efforts. Finally, assume that each individual i's utility function equals $z_i - ca_i$, where z_i is the amount of money received by i.

Radner et al. (1986) introduced this partnership in the context of repeated games. They considered the problem of providing incentives for everyone to work—if not all the time at least most of the time—without needing to inject or withdraw resources from the group as a whole. They effectively showed that in this environment there do not exist output-contingent rewards that both (i) balance the group's budget, i.e., the sum of individual payments always equals zero, and (ii) induce everyone to work most of the time, let alone all of the time. Indeed, for everyone to work at all they must be rewarded when output is good. However, this arrangement violates budget balance, since everyone being rewarded when output is good clearly implies that the sum of payments across agents is greater when output is good than when it is bad.

An arrangement that still does not solve the partnership problem but nevertheless induces most people to work is appointing an agent to play the role of Holmström's principal. Call this agent 1 and define output-contingent payments to individuals as follows. For $i = 2, \ldots, n$, let $\zeta_i(g) = \overline{z}$ and $\zeta_i(b) = 0$ be agent i's output-contingent money payment, for some $\overline{z} \geq 0$. To satisfy budget balance, agent 1's transfer equals

$$\zeta_1 \;=\; -\sum_{i=2}^{n} \zeta_i.$$

By construction, the budget is balanced. It is easy to see that everyone but agent 1 will work if $\overline{z}$ is sufficiently large. However, agent 1 has the incentive to shirk.[7]

[7]This contract follows Holmström's suggestion to the letter: agent 1 is a "fixed" principal who absorbs the incentive payments to all others by "breaking" everyone else's budget constraint.

With mediated contracts it is possible to induce everyone to work most of the time. Indeed, consider the following incentive scheme. For any small $\varepsilon > 0$, a mediator or machine asks every individual to work (call this event $\mathbf{1}$) with probability $1-\varepsilon$. With probability ε/n, agent i is picked (assume everyone is picked with equal probability) and secretly asked to shirk, while all others are asked to work (call this event $\mathbf{1}_{-i}$). For $i = 1, \ldots, n$, let $\zeta_i(g|\mathbf{1}) = \zeta_i(b|\mathbf{1}) = 0$ be agent i's contingent transfer if the mediator asked everyone to work. Otherwise, if agent i was secretly asked to shirk, for $j \neq i$ let $\zeta_j(g|\mathbf{1}_{-i}) = \overline{z}$ and $\zeta_j(b|\mathbf{1}_{-i}) = 0$ be agent j's transfer. For agent i, let

$$\zeta_i(\mathbf{1}_{-i}) \;=\; -\sum_{j\neq i} \zeta_j(\mathbf{1}_{-i}).$$

By construction, this contract is budget-balanced. It is also incentive compatible. Indeed, it is clear that asking an agent to shirk is always incentive compatible. If agent i is recommended to work, incentive compatibility requires that

$$\tfrac{\varepsilon}{n}(n-1)P(n-1)\overline{z} - c \;\geq\; \tfrac{\varepsilon}{n}(n-1)P(n-2)\overline{z},$$

which is satisfied if $\overline{z}$ is sufficiently large because P is strictly increasing. Under this contract, everyone works with probability $1-\varepsilon$, for any $\varepsilon > 0$, by choosing $\overline{z}$ appropriately, so everyone working is approximated with budget balanced transfers.

The arrangement above solves the partnership problem of Radner et al. (1986) by occasionally appointing a secret principal, who is asked to shirk and is used to restore budget balance while providing incentives for all others. In order to induce everyone to work, mediated contracts allow for different principals to be used for different workers. The fact that these principals are appointed secretly makes it possible for them to be effectively used simultaneously. Finally, they are chosen only seldom to reduce the inherent loss from having a principal in the first place.

Example 1 reveals the logic behind our main result, Theorem 1. If a worker deviates (i.e., shirks) then he will decrease the probability of g not only when everyone else is asked to work, but also when a principal is appointed. In this latter case, innocence can be attributed to the principal, so the deviator can be punished by having every worker pay the principal. In other words, for each worker and any deviation by the worker there is a profile of actions by others such that his deviation can be statistically distinguished from someone else's (in this case, a principal, since the principal's deviation would raise the probability of g). This turns out to be not only necessary but also sufficient for solving any partnership problem.

3 Model

This section develops our model of mediated partnerships. It describes the environment, the timing of agents' interaction, notions of enforcement, and attribution.

Let $I = \{1, \ldots, n\}$ be a finite set of agents, A_i a finite set of actions available to any agent $i \in I$, and $A = \prod_i A_i$ the (nonempty) space of action profiles. Write $v : I \times A \to \mathbb{R}$ for the profile of agents' utility functions, where $v_i(a)$ denotes the utility to any agent $i \in I$ from any action profile $a \in A$. A *correlated strategy* is any probability measure $\sigma \in \Delta(A)$.[8] Let S_i be a finite set of *private signals* observable only by agent $i \in I$ and S_0 a finite set of *publicly verifiable* signals. Let $S := \prod_{j=0}^n S_j$ be the (nonempty) space of all signal profiles. A *monitoring technology* is a measure-valued map $\Pr : A \to \Delta(S)$, where $\Pr(s|a)$ denotes the conditional probability that signal profile s was observed given that action profile a was played.

Incentives are provided to agents with linear transfers. An *incentive scheme* is any map $\zeta : I \times A \times S \to \mathbb{R}$ that assigns monetary payments contingent on individuals, recommended actions, and reported signals, all of which are assumed verifiable.

Definition 1. A *contract* is any pair (σ, ζ), where σ is a correlated strategy and ζ is an incentive scheme. It is called *standard* if $\zeta : I \times S \to \mathbb{R}$, i.e., payments do not depend on recommendations. Otherwise, the contract is called *mediated*.

Standard contracts are a special case of mediated ones but not otherwise. For instance, the secret principal of Section 2 is a nonstandard mediated contract, since payments depend on recommendations. The literature has mostly focused on standard contracts to study incentives, whereas this paper concentrates on mediated ones.

The timing of agents' interaction unfolds as follows. Firstly, agents agree on some contract (σ, ζ). A profile of recommendations is drawn according to σ and made to agents confidentially by some mediator. Agents then simultaneously take some action, which is neither verifiable nor directly observable. Next, agents observe unverifiable private signals and submit a verifiable report of their observations before observing the public signal (the timing of signals is not essential, just simplifying). Finally, recommendation- and report-contingent transfers are made according to ζ.

[8] If X is a finite set, $\Delta(X) = \{\mu \in \mathbb{R}^X_+ : \sum_x \mu(x) = 1\}$ is the set of probability vectors on X.

If every agent is honest and obedient, i's ex ante expected utility from (σ, ζ) is

$$\sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a,s) \Pr(s|a).$$

Of course, agent i may disobey his recommendation a_i to play some other action b_i and lie about his privately observed signal. A *reporting strategy* is a map $\rho_i : S_i \to S_i$, where $\rho_i(s_i)$ is the reported signal when i privately observes s_i. For instance, the truthful reporting strategy is the identity map $\tau_i : S_i \to S_i$ with $\tau_i(s_i) = s_i$. Let R_i be the set of all reporting strategies for agent i. The space of *deviations* for i is $A_i \times R_i$. For every agent i and every deviation (b_i, ρ_i), the conditional probability that $s \in S$ will be reported when everyone else is honest and plays $a_{-i} \in A_{-i}$ equals[9]

$$\Pr(s|a_{-i}, b_i, \rho_i) \ := \sum_{t_i \in \rho_i^{-1}(s_i)} \Pr(s_{-i}, t_i | a_{-i}, b_i).$$

A contract (σ, ζ) is *incentive compatible* if obeying recommendations and reporting honestly is optimal for every agent when everyone else is honest and obedient, i.e.,

$$\forall i \in I, a_i \in A_i, (b_i, \rho_i) \in A_i \times R_i,$$
$$\sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) \ \leq \sum_{(a_{-i},s)} \sigma(a) \zeta_i(a,s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)). \quad (*)$$

The left-hand side above reflects the *utility gain*[10] for an agent i from playing b_i when asked to play a_i. The right-hand side reflects his *monetary loss* from deviating to (b_i, ρ_i) relative to honesty and obedience. Such a loss originates from two sources. On the one hand, playing b_i instead of a_i may change conditional probabilities over signals. On the other, reporting according to ρ_i may affect conditional payments.

Definition 2. A correlated strategy σ is *exactly enforceable* (or simply *enforceable*) if there is an incentive scheme $\zeta : I \times A \times S \to \mathbb{R}$ to satisfy $(*)$ for all (i, a_i, b_i, ρ_i) and

$$\forall (a,s), \qquad \sum_{i \in I} \zeta_i(a,s) \ = \ 0. \qquad (**)$$

Call σ *virtually enforceable* if there exists a sequence $\{\sigma^m\}$ of enforceable correlated strategies such that $\sigma^m \to \sigma$.

A correlated strategy is enforceable if it is both incentive compatible and budget-balanced. It is virtually enforceable if it is the limit of enforceable ones. This requires budget balance along the way, not just asymptotically. For instance, in Example 1, everybody working is virtually enforceable, but not exactly enforceable.

[9]We use the notation $s = (s_{-i}, s_i)$ for $s_i \in S_i$ and $s_{-i} \in S_{-i} = \prod_{j \neq i} S_j$; similarly for $a = (a_{-i}, a_i)$.
[10]Specifically, probability-weighted utility, weighted by $\sigma(a_i) = \sum_{a_{-i}} \sigma(a)$, the probability of a_i.

We end this section by defining a key condition called identifying obedient players, which will be shown to characterize enforcement. We begin with some preliminaries.

A *deviation plan* for any agent i is a map $\alpha_i : A_i \to \Delta(A_i \times R_i)$, where $\alpha_i(b_i, \rho_i|a_i)$ stands for the probability that i deviates to (b_i, ρ_i) when recommended to play a_i. For any σ and any α_i, let $\Pr(\sigma, \alpha_i) \in \Delta(S)$, defined pointwise by

$$\Pr(s|\sigma, \alpha_i) \;=\; \sum_{a \in A} \sigma(a) \sum_{(b_i, \rho_i)} \Pr(s|a_{-i}, b_i, \rho_i)\alpha_i(b_i, \rho_i|a_i),$$

be the vector of report probabilities if agent i deviates from σ according to α_i.

Definition 3. A profile of deviation plans $\alpha = (\alpha_1, \ldots, \alpha_n)$ is *unattributable* if

$$\forall a \in A, \qquad \Pr(a, \alpha_1) \;=\; \cdots \;=\; \Pr(a, \alpha_n).$$ [11]

Call α *attributable* if it is not unattributable, i.e., there exist agents i and j such that $\Pr(a, \alpha_i) \neq \Pr(a, \alpha_j)$ for some $a \in A$.

Intuitively, a profile of deviation plans is unattributable if any unilateral deviation in the profile would lead to the same distribution over reports. Heuristically, after a deviation plan belonging to some unattributable profile, even if the fact that someone deviated was detected, anyone could have been the culprit.

Call α_i *disobedient* if $\alpha_i(b_i, \rho_i|a_i) > 0$ for some $a_i \neq b_i$, i.e., it disobeys some recommendation with positive probability. A disobedient deviation plan may be "honest," i.e., ρ_i may equal τ_i. However, dishonesty by itself (obeying recommendations but choosing $\rho_i \neq \tau_i$) is not labeled as disobedience. A *disobedient profile* of deviation plans is any $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that α_i is disobedient for at least one agent i.

Definition 4. A monitoring technology *identifies obedient agents* (IOA) if every disobedient profile of deviation plans is attributable.

IOA means that for every disobedience by some arbitrary agent i and every profile of others' deviation plans, an action profile exists such that i's unilateral deviation has a different effect on report probabilities from at least one other agent. For instance, the monitoring technology of Example 1 identifies obedient agents. There, if a worker shirks then good news becomes *less* likely, whereas if a principal works then good news becomes *more* likely. Hence, a profile of deviation plans with i disobeying is attributable by just having another agent behave differently from i. This implies IOA. Intuitively, IOA holds by using different principals for different workers.

[11]We slightly abuse notation by identifying action profiles with pure correlated strategies.

4 Results

This section presents the paper's main results, characterizing enforceable outcomes in terms of the monitoring technology, with and without mediated contracts. We begin with a key lemma that provides a dual characterization of IOA.

Lemma 1. *A monitoring technology identifies obedient agents if and only if there exists a function $\xi : I \times A \times S \to \mathbb{R}$ such that $\sum_i \xi_i(a,s) = 0$ for every (a,s) and*

$$\forall(i, a_i, b_i, \rho_i), \qquad 0 \ \le \sum_{(a_{-i},s)} \xi_i(a,s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

with a strict inequality *whenever $a_i \neq b_i$.*

Intuitively, Lemma 1 shows that IOA is equivalent to the existence of budget-balanced "probability-weighted" transfers ξ such that (i) the budget is balanced, (ii) no deviation is profitable, and (iii) every disobedience incurs a strictly positive monetary cost. If every action profile is recommended with positive probability, i.e., if $\sigma \in \Delta^0(A) := \{\sigma \in \Delta(A) : \sigma(a) > 0 \ \forall a \in A\}$ is any completely mixed correlated strategy, then there is an incentive scheme ζ with $\xi_i(a,s) = \sigma(a)\zeta_i(a,s)$ for all (i,a,s). Therefore, IOA implies that given $\sigma \in \Delta^0(A)$ and ξ satisfying Lemma 1, we may choose ζ appropriately to overcome all incentive constraints simultaneously. Hence, the second half of Lemma 1 is equivalent to every completely mixed correlated strategy being exactly enforceable. Approximating each correlated strategy with completely mixed ones establishes our main result, Theorem 1 below.

Theorem 1. *A monitoring technology identifies obedient agents if and only if for any profile of utility functions, every correlated strategy is virtually enforceable.*

Theorem 1 characterizes monitoring technologies such that "everything" is virtually enforceable, regardless of preferences. It says that identifying obedient agents in a weak sense is not only necessary but also sufficient for virtual enforcement. Intuitively, if after a deviation some innocent agent can be statistically identified then that agent can be rewarded at the expense of everyone else, thereby punishing the deviator. Heuristically, if a profile of deviation plans can be attributed then there is an incentive scheme that discourages every deviation in that profile. Theorem 1 says that for every disobedient profile of deviation plans there is a scheme that discourages it if and only if there is a scheme that discourages all profiles of deviation plans simultaneously.

To put Theorem 1 in perspective, consider the scope of enforcement with standard contracts. By Example 1, IOA is generally not enough for enforcement with standard contracts, but the following strengthening is. Given a subset $B \subset A$ of action profiles and an agent i, let $B_i := \{b_i \in A_i : \exists b_{-i} \in A_{-i} \text{ s.t. } b \in B\}$ be the projection of B on A_i. Call a deviation plan α_i *B-disobedient* if it is disobedient at some $a_i \in B_i$, i.e., if $\alpha_i(b_i, \rho_i | a_i) > 0$ for some $b_i \neq a_i \in B_i$. A *B-disobedient profile* of deviation plans is any $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that α_i is B-disobedient for some agent i. Given $\sigma \in \Delta(A)$, α is *attributable at σ* if there exist agents i and j such that $\Pr(\sigma, \alpha_i) \neq \Pr(\sigma, \alpha_j)$, and say Pr *identifies obedient agents at σ* (IOA-σ) if every supp σ-disobedient[12] profile of deviation plans is attributable at σ. Intuitively, IOA-σ differs from IOA in that IOA allows for different α's to be attributed at different σ's, whereas IOA-σ does not.

Theorem 2. *A monitoring technology identifies obedient agents at σ if and only if for any profile of utility functions, σ is exactly enforceable with a standard contract.*

Theorem 2 characterizes enforceability with standard contracts of any correlated strategy σ in terms of IOA-σ. Intuitively, it says that enforcement with standard contracts requires that every α be attributable at the same σ.[13] Theorem 2 also sheds light into the value of mediated contracts. Indeed, the proof of Theorem 1 shows that enforcing a completely mixed correlated strategy (i.e., such that $\sigma(a) > 0$ for all a) only requires IOA, by allowing for different profiles of deviation plans to be attributable at different action profiles. This condition is clearly weaker than IOA-σ. On the other hand, IOA is generally not enough to enforce a given pure-strategy profile a, as Example 1 shows with $a = \mathbf{1}$ there. Since agents receive only one recommendation under a, there is no use for mediated contracts, so by Theorem 2 IOA-a characterizes exact enforcement of a with both standard and mediated contracts.[14]

Now consider the intermediate case where σ has arbitrary support. Fix a subset of action profiles $B \subset A$. A profile of deviation plans $\alpha = (\alpha_1, \ldots, \alpha_n)$ is *B-attributable* if there exist agents i and j such that $\Pr(a, \alpha_i) \neq \Pr(a, \alpha_j)$ for some $a \in B$. Otherwise, α is called *B-unattributable*. For instance, A-attribution is just attribution. Say Pr *B-identifies obedient agents* (B-IOA) if every B-disobedient deviation plan is B-attributable. For instance, A-IOA is just IOA, and $\{a\}$-IOA equals IOA-a.

[12] By definition, supp $\sigma = \{a \in A : \sigma(a) > 0\}$ is the support of σ.

[13] Even for virtual enforcement with standard contracts the same σ must attribute all α's. E.g., in Example 1 there is no sequence $\{\sigma^m\}$ with $\sigma^m \to \mathbf{1}$ and Pr satisfying IOA-σ^m for all m.

[14] Again, we abuse notation by labeling a as both an action profile and a pure correlated strategy.

Theorem 3. *For any subset $B \subset A$, the following are equivalent: (1) The monitoring technology B-identifies obedient agents. (2) Every correlated strategy with support equal to B is enforceable for any profile of utility functions. (3) Some fixed correlated strategy with support equal to B is enforceable for any profile of utility functions.*

Theorem 3 characterizes enforcement with mediated contracts of any correlated strategy σ with supp σ-IOA. Hence, only the support of a correlated strategy matters for its enforcement for all preferences. Moreover, any other correlated strategy with support contained in supp σ becomes virtually enforceable, just as with Theorem 1. Intuitively, mediated contracts allow for different actions in the support of a correlated strategy to attribute different deviation profiles, unlike standard contracts, as shown above. Therefore, clearly IOA-σ is more restrictive than supp σ-IOA.

Although the results above focused on enforcement for all utility profiles, restricting attention to fixed preferences does not introduce additional complications and yields similar results. Indeed, fix a profile $v : I \times A \to \mathbb{R}$ of utility functions. A natural weakening of IOA involves allowing unprofitable profiles of deviation plans to be unattributable. A profile of deviation plans α is called *σ-profitable* if

$$\sum_{(i,a,b_i,\rho_i)} \sigma(a)\alpha_i(b_i,\rho_i|a_i)(v_i(a_{-i},b_i) - v_i(a)) \ > \ 0.$$

Intuitively, the profile α is σ-profitable if the sum of utility gains from each unilateral deviation plan in the profile is positive. Enforcement now amounts to the following.

Theorem 4. *(1) Every σ-profitable profile of deviation plans is supp σ-attributable if and only if σ is enforceable. (2) Every σ-profitable profile of deviation plans is attributable at σ if and only if σ is enforceable with a standard contract.*

Theorem 4 characterizes enforceability with and without mediated contracts. It describes how mediated contracts add value by relaxing the burden of attribution: Every profile α that is attributable at σ is supp σ-attributable, but not conversely. For instance, in Example 1, let $\sigma(S)$ be the probability that $S \subset I$ are asked to work, and suppose that $\sigma(I) > 0$. Let α be the deviation profile where every agent i shirks with probability p_i if asked to work (and obeys if asked to shirk), with $p_i = \sigma(I)[P(n-1) - P(n)]/\sum_{S \ni i} \sigma(S)[P(|S|-1) - P(|S|)] \in (0,1]$. By construction, the probability of good output equals $\sigma(I)P(n-1) + \sum_{S \neq I} \sigma(S)P(|S|)$, which is independent of i. Therefore, α is not attributable at any σ with $\sigma(I) > 0$. However, α is attributable, since the monitoring technology identifies obedient agents.

5 Discussion

In this section we decompose IOA to understand it better assuming public monitoring, establish its genericity, and add participation and liability constraints.

5.1 Public Monitoring

To help understand IOA, let us temporarily restrict attention to *publicly verifiable* monitoring technologies, i.e., such that $|S_i| = 1$ for all $i \neq 0$. In this case, IOA can be naturally decomposed into two parts. We formalize this decomposition next.

A deviation plan α_i for any agent i is *detectable* if $\Pr(a, \alpha_i) \neq \Pr(a)$ at some $a \in A$. Say Pr *detects unilateral disobedience* (DUD) if every disobedient deviation plan is detectable,[15] where different action profiles may be used to detect different deviation plans. Say *detection implies attribution* (DIA) if for every detectable deviation plan α_i and every deviation profile α_{-i}, $\alpha = (\alpha_{-i}, \alpha_i)$ is attributable. Intuitively, DIA says that if a deviation is detected, someone can be (statistically) ruled out as innocent.

Theorem 5. *A publicly verifiable monitoring technology identifies obedient agents if and only if* (i) *it detects unilateral disobedience and* (ii) *detection implies attribution.*

An immediate example of DIA is Holmström's (1982) principal, i.e., an individual i_0 with no actions to take or signals to observe (both A_{i_0} and S_{i_0} are singletons). The principal is automatically obedient, so every detectable deviation plan can be discouraged with budget balance by rewarding him and punishing everyone else. DIA isolates this idea and finds when the principal's role can be fulfilled internally. It helps to provide budget-balanced incentives by identifying innocent individuals to be rewarded and punishing all others (if necessary) when a disobedience is detected.

Next, we give a dual characterization of DIA that sheds light into the role it plays in Theorem 1. A publicly verifiable monitoring technology Pr *clears every budget* (CEB) if for every $K : A \times S \to \mathbb{R}$ there exists $\xi : I \times A \times S \to \mathbb{R}$ such that

$$\forall (a, s), \qquad \sum_{i \in I} \xi_i(a, s) = K(a, s), \quad \text{and}$$

$$\forall (i, a_i, b_i), \qquad 0 \leq \sum_{(a_{-i}, s)} \xi_i(a, s)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)).$$

[15] This condition on a monitoring technology was introduced and analyzed by Rahman (2008).

The function $K(a, s)$ may be regarded as a budgetary surplus or deficit for each combination of recommended action and realized signal. Intuitively, CEB means that any budget can be attained by some payment scheme that avoids disrupting any incentive compatibility constraints. As it turns out, this is equivalent to DIA.

Theorem 6. *A publicly verifiable monitoring technology clears every budget if and only if detection implies attribution.*

This result helps clarify the roles of DUD and DIA in Theorem 1. Rahman (2008) shows that DUD characterizes virtual enforcement without budget balance of any correlated strategy σ, regardless of preferences. CEB guarantees the existence of a further contract to absorb any budgetary deficit or surplus of the original contract without violating any incentive constraints. Therefore, the original contract plus this further contract can now virtually enforce σ with a balanced budget.[16]

If the monitoring technology is not publicly verifiable, the above decomposition of IOA does not emerge naturally. It is easy to see that DIA plus DUD is sufficient for IOA. However, necessity fails in general because there may exist dishonest but otherwise obedient deviations that do not directly affect anyone's utility, and as such IOA allows them to be unattributable even if detectable. With verifiability, every deviation may in principle affect agents directly. The next example makes this point.

Example 2. There are three agents and A_i is a singleton for every agent i, so IOA is automatically satisfied. There are no public signals and each agent observes a binary private signal: $S_i = \{0, 1\}$ for all i. The monitoring technology is

$$\Pr(s) := \begin{cases} \frac{6}{25} & \text{if } \sum_i s_i = 3 \\ \frac{3}{25} & \text{if } \sum_i s_i = 1 \text{ or } 2 \\ \frac{1}{25} & \text{if } \sum_i s_i = 0 \end{cases}$$

The following is a profile of (trivially obedient) unattributable deviation plans that are also detectable, violating DIA. Suppose that agent i deviates by lying with probability $2/5$ after observing $s_i = 1$ and lying with probability $3/5$ after observing $s_i = 0$. For every agent i, the joint distribution of reported private signals becomes:

$$\Pr(s) = \begin{cases} \frac{27}{125} & \text{if } \sum_i s_i = 3 \\ \frac{18}{125} & \text{if } \sum_i s_i = 2 \\ \frac{12}{125} & \text{if } \sum_i s_i = 1 \\ \frac{8}{125} & \text{if } \sum_i s_i = 0 \end{cases}$$

[16] A comparable argument is provided by d'Aspremont et al. (2004) for Bayesian mechanisms.

5.2 Genericity

Genericity of IOA is discussed next. To motivate, consider firstly a negative result.

Theorem 7. *Identifying obedient agents is impossible with only two agents, at least two actions per agent and no public information.*

Theorem 7 simply says that with two agents and no public signals it is always possible to blame the other agent for a deviation. Since it is impossible to identify who deviated, by elimination it is also impossible to identify who did not deviate. Fortunately, IOA almost always holds beyond this environment. To show this, relabel the agents so that $i < j$ if $|S_i| \leq |S_j|$, and let k be the number of agents i with $|S_i| = 1$, i.e., not sending reports. Without loss, assume $i < j$ if $|A_i| \leq |A_j|$ for all $i, j \leq k$.

Theorem 8. *IOA is generic if for every agent i, (a) $|A_i| - 1 \leq |A_{-i}| (|S_{-i}| - 1)$ when $|S_i| = 1$, (b) $|A_i| (|S_i| - 1) \leq |A_{-i}| - 1$ when $|S_{-i}| = 1$, and (c) $|A_i| |S_i| \leq |A_{-i}| |S_{-i}|$ when both $|S_i| > 1$ and $|S_{-i}| > 1$, as well as*

$$\sum_{i=1}^{n} (|A_i| |S_i|)^2 - 1 - \chi_n |A_n| |S_n| (|A_n| - 1) \leq (n-1) |A| |S| - (k-1)(|A| - |A_k| + 1) + \sum_{i=1}^{k-1} |A_i| ,$$

where $\chi_n = 1$ if $|S_{-n}| = 1$ and 0 otherwise, and agents are ordered as above.

Conditions (a,b,c) above imply that DUD is generic, and the last condition that DIA is generic. To explain Theorem 8, consider some examples. Firstly, if agent 1 is a principal, i.e., $|A_1| |S_1| = 1$, then IOA is generic if (a,b,c) hold, so DUD is generic.

Example 3. If every agent has the same number of actions, so $|A_i| = m$ for all i, and signals are verifiable, so $|S| = |S_0| = \ell$, then IOA is generic if $m - 1 \leq m^{n-1}(\ell - 1)$ and $nm^2 - 1 \leq (n-1)[m^n(\ell - 1) + 2m - 1]$, which holds for all $\ell > 1$ and $m \geq 1$ if $n > 2$. Hence, IOA is generic with at least three agents and two public signals, and two agents with at least as many signals as actions per player, i.e., $\ell \geq m$.

Example 4. If $|A_i| = m$, $|S_i| = \ell$, and $|S_0| = 1$, i.e., there are no public signals, then IOA is generic when $\ell > 1$, $m\ell \leq m^{n-1}\ell^{n-1}$, and $nm^2\ell^2 - 1 \leq (n-1)m^n\ell^n$, which holds for all $\ell, m > 1$ and $n > 2$. Hence, IOA is generic by adding a similar agent to Theorem 7 even without public information.

Example 5. If $|A_i| = m$ and $|S| = |S_n| = 2$, so only agent n observes a (binary) signal, then IOA is generic when $m - 1 \leq m^{n-1}$, $m \leq m^{n-1} - 1$ and $(n+1)m^2 \leq nm^n + (n-3)(2m-1)$. All inequalities hold if $m > 1$ and $n > 2$. Hence, IOA is generic if only one agent observes a binary signal and there are at least three agents.

5.3 Participation and Liability

Individual rationality—or participation—constraints are easily incorporated into the present study of incentives, by imposing the following family of inequalities:

$$\forall i \in I, \qquad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a,s) \Pr(s|a) \ \geq \ 0.$$

Theorem 9. *Participation is not a binding constraint if $\sum_i v_i(a) \geq 0$ for all $a \in A$.*

Theorem 9 generalizes standard results (e.g., d'Aspremont and Gérard-Varet, 1998, Lemma 1) to our setting.

Next, we study limited liability given $z \in \mathbb{R}^I_+$, by imposing constraints of the form $\zeta_i(a,s) \geq -z_i$. Intuitively, an agent can never pay any more than z_i. Call z_i agent i's *liability*, and z the *distribution of liability*. A group's *total liability* is defined by $\widehat{z} = \sum_i z_i$. Without participation constraints, Theorem 5 of Legros and Matsushima (1991) and Theorem 4 of Legros and Matthews (1993) easily generalize to this setting.

Theorem 10. *In the absence of participation constraints, only total liability affects the set of enforceable outcomes, not the distribution of liability.*

Including participation constraints leads to the following characterization.

Theorem 11. *The correlated strategy σ is enforceable with individual rationality and liability limited by z if and only if*

$$\sum_{(a,i,b_i,\rho_i)} \sigma(a) \alpha_i(b_i, \rho_i | a_i)(v_i(a_{-i}, b_i) - v_i(a)) \ \leq \ \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \sum_{i \in I} z_i$$

for every (α, π) such that α is a profile of deviation plans and $\pi = (\pi_1, \ldots, \pi_n) \geq 0$, where $\widehat{\eta} := \sum_{(a,s)} \min_i \{\Pr(s|a,\alpha_i) - (1+\pi_i)\Pr(s|a)\}$ and $v_i(\sigma) = \sum_a \sigma(a) v_i(a)$.

Theorem 11 generalizes Theorems 9 and 10, as the next result shows.

Corollary 1. *Suppose that σ is enforceable with individual rationality and liability limited by z.* (i) *If $v_i(\sigma) \geq z_i$ then agent i's participation is not a binding constraint.* (ii) *The distribution of liability does not matter within the subset t of agents whose participation constraint is not binding, i.e., σ is also enforceable with individual rationality and liability limited by any z' with $z_j = z'_j$ for $j \in I \setminus t$ and $\sum_{i \in t} z_i = \sum_{i \in t} z'_i$.*

6 Literature

To help identify this paper's contribution, let us now compare its results with the literature. Broadly, the paper contributes: (i) a systematic analysis of partnerships that fully exploit internal communication, and (ii) results showing that attribution and IOA yield the weakest requirements on a monitoring technology for enforcement and virtual enforcement. IOA enjoys the key property that different action profiles can be used to attribute different profiles of disobedient deviation plans, in contrast with the literature, which we discuss below.

In contract theory, Legros and Matsushima (1991) characterize exact enforcement with standard contracts and publicly verifiable signals, but they do not interpret their results in terms of attribution, nor do they consider virtual enforcement. Another related paper is d'Aspremont and Gérard-Varet (1998). In the same context as Legros and Matsushima (1991), they derive intuitive sufficient conditions for enforcement. A closer paper to ours is Legros and Matthews (1993), who study virtual enforcement with standard contracts and deterministic output. They propose a contract that uses mixed strategies to identify non-shirkers whenever possible,[17] but the same strategies must identify non-shirkers after every deviation, unlike mediated contracts. Their contract fails to provide the right incentives if output given efforts is stochastic and its distribution does not have a "moving support," i.e., the support does not depend on efforts. The key difference between their contract and ours is that mediated partnerships correlate agents' payoffs not just to output, but also to others' mixed strategies. As a result, mediated partnerships can virtually enforce efficient behavior even without a moving support, as Example 1 and Theorem 1 show.

In the context of repeated games, the closest papers to ours may be Kandori (2003), Aoyagi (2005) and Tomala (2005). They establish versions of the Folk Theorem by interpreting players' continuation values as linear transfers. Kandori allows agents to play mixed strategies and report on the realization of such mixtures after observing a public signal. He considers contracts contingent on the signals and these reports. Although his contracts are nonstandard, they fail to fully employ communication. For instance, they fail to provide incentives in Example 1. Aoyagi uses dynamic mediated strategies that rely on "ε-perfect" monitoring, and fail if monitoring is costly or one-sided. Finally, Tomala studies a class of recursive communication equilibria.

[17] Identifying non-shirkers was suggested in mechanism design by Kosenok and Severinov (2008).

There are several differences between these papers and ours. One especially noteworthy difference is that to prove the Folk Theorem they make much more restrictive assumptions than IOA, structurally similar to *pairwise full rank* (PFR) of Fudenberg et al. (1994). Intuitively, PFR-like conditions ask to identify deviators instead of just non-deviators. To see this, let us focus for simplicity on public monitoring and recall the decomposition of IOA into DUD and DIA (Theorem 5).

For every i, let C_i (called the *cone* of agent i) be the set of all $\eta \in \mathbb{R}^{A\times S}$ with

$$\forall(a,s), \qquad \eta(a,s) \;=\; \sum_{b_i\in A_i} \alpha_i(b_i|a_i)(\Pr(s|a_{-i},b_i) - \Pr(s|a))$$

for some $\alpha_i : A_i \to \Delta(A_i)$. DIA imposes the following restriction on agents' cones:

$$\bigcap_{i\in I} C_i \;=\; \{\mathbf{0}\},$$

where $\mathbf{0}$ stands for the origin of $\mathbb{R}^{A\times S}$.

In other words, agents' cones do not overlap. PFR implies that for *every pair* of agents, their cones do not overlap. Intuitively, this means that upon any deviation it is possible to identify the deviator's identity. On the other hand, DIA only requires that *all* agents' cones fail to overlap simultaneously. Thus, it is possible to provide budget-balanced incentives even if there are two agents whose cones overlap (i.e., their intersection is larger than just the origin), so PFR fails. In general, DIA does not even require that there exist two agents whose cones fail to overlap, in contrast with *local compatibility* of d'Aspremont and Gérard-Varet (1998). Figure 1 below illustrates this point.[18]

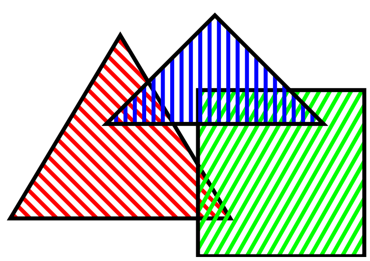

Figure 1: A cross-section of three non-overlapping cones in $\mathbb{R}^3$ (pointed at the origin behind the page) such that every pair of cones overlaps.

[18] Figure 1 is not pathological. Indeed, Example 1 may be viewed as a version of Figure 1.

7 Conclusion

Mediated partnerships embrace the idea that—as part of an economic organization—it may be beneficial for private information to be allocated differently across agents to provide the right incentives. As Example 1 illustrates, mediated partnerships can enforce outcomes that standard ones simply cannot. Indeed, mediated contracts can provide the right incentives in partnerships with stochastic output whose distribution fails to exhibit a "moving support" (i.e., the support is independent of effort), even without complementarities in production. Standard contracts cannot.

In general, mediated partnerships are enforceable if and only if it is possible to identify obedient agents. This means that after any unilateral deviation, innocence is statistically attributable to someone, although different actions may be used to attribute innocence after different deviations. Informationally, this is clearly less costly than attempting to attribute guilt, as well as using the same actions to attribute innocence after every deviation. This latter difference exactly captures the value of mediated partnerships.

A Proofs

Lemma 1. By the Theorem of the Alternative (Rockafellar, 1970, p. 198), the condition on ξ above fails if and only if $\lambda \geq 0$ and $\eta \in \mathbb{R}^{A\times S}$ exist with $\lambda_i(a_i, b_i, \rho_i) > 0$ for some (i, a_i, b_i, ρ_i) with $a_i \neq b_i$ and

$$\forall (i, a, s), \qquad \sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) \;=\; \eta(a, s),$$

where η is independent of i. Let $\Lambda = \max_{(i,a_i)} \sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) > 0$. Define

$$\alpha_i(b_i, \rho_i|a_i) := \begin{cases} \lambda_i(a_i, b_i, \rho_i)/\Lambda & \text{if } (b_i, \rho_i) \neq (a_i, \tau_i), \text{ and} \\ 1 - \sum_{(b_i,\rho_i)\neq(a_i,\tau_i)} \lambda_i(a_i, b_i, \rho_i)/\Lambda & \text{otherwise.} \end{cases}$$

By construction, α_i is disobedient and unattributable (using α_{-i}): IOA fails. □

Theorem 1. Follows from the paragraph preceding the statement of the theorem. □

Theorem 2. By the Alternative Theorem, Pr satisfies IOA-σ if and only if there is a signal-contingent scheme $\zeta : I \times S \to \mathbb{R}$ such that $\sum_i \zeta_i(s) = 0$ for all s and

$$\forall i \in I, a_i \in B_i, (b_i, \rho_i) \in A_i \times R_i, \qquad 0 \;\leq\; \sum_{(a_{-i},s)} \sigma(a)\zeta_i(s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)),$$

with a strict inequality if $a_i \neq b_i$, where $B_i = \{a_i \in A_i : \exists a_{-i} \text{ s.t. } \sigma(a) > 0\}$. Call this dual condition IOA*-σ. By scaling ζ as necessary, IOA*-σ clearly implies that any deviation gains can be outweighed by monetary losses. Conversely, if IOA-σ fails then there is a profile of deviation plans α such that $\Pr(\sigma, \alpha_i) = \Pr(\sigma, \alpha_j)$ for all (i, j) and there is an agent i^* such that α_{i^*} satisfies $\alpha_{i^*}(b_{i^*}, \rho_{i^*}|a_{i^*}) > 0$ for some $a_{i^*} \in B_{i^*}$, and $b_{i^*} \neq a_{i^*}$. For all a_{-i^*}, let $0 = v_{i^*}(a) < v_{i^*}(a_{-i^*}, b_{i^*}) = 1$ and $v_j(a) = v_j(a_{-i^*}, b_{i^*}) = 0$ for all $j \neq i^*$. Now σ cannot be enforced by any $\zeta : I \times S \to \mathbb{R}$ such that $\sum_i \zeta_i(s) = 0$ for all s, since $\sum_{(i,b_i,\rho_i)} \alpha_i(b_i, \rho_i|a_i) \sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) > \sum_{(i,s)} \zeta_i(s)(\Pr(s|\sigma, \alpha_i) - \Pr(s|\sigma)) = 0$, being a nonnegative linear combination of incentive constraints, will violate at least one. □

Theorem 3. "(1) ⇔ (2)" follows by applying a version of the proof of Lemma 1 and Theorem 1 after replacing B with A. "(1) ⇔ (3)" follows similarly, after fixing any correlated strategy σ with support equal to B. □

Theorem 4. (1) follows by applying the proof of Lemma 1 with both σ and v fixed to the incentive compatibility constraints (∗). (2) follows by a similar version of the proof of Theorem 2, again with both σ and v fixed. □

Theorem 5. IOA clearly implies DUD (just replace α_{-i} with honesty and obedience for every α_i in the definition of attribution). By IOA, if a profile α is unattributable then it is obedient, hence every deviation plan in the profile is undetectable (since the monitoring technology is publicly verifiable), and DIA follows. Conversely, DIA implies that every unattributable α_i is undetectable, and by DUD, every undetectable α_i is obedient. □

Theorem 6. Consider the following primal problem: Find a feasible ξ to solve

$$\forall(i, a_i, b_i),\ 0 \leq \sum_{(a_{-i},s)} \xi_i(a, s)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)),\ \text{and } \forall(a, s),\ \sum_{i \in I} \xi_i(a, s) = K(a, s).$$

The dual of this problem is given by

$$\inf_{\lambda \geq 0, \eta} \sum_{(a,s)} \eta(a, s)K(a, s) \ \text{ s.t. } \forall(i, a, s),\ \sum_{b_i \in A_i} \lambda_i(a_i, b_i)(\Pr(s|a_{-i}, b_i) - \Pr(s|a)) \ = \ \eta(a, s).$$

If CEB is satisfied, then the value of the primal equals 0 for any $K : A \times S \to \mathbb{R}$. By the Strong Duality Theorem, the value of the dual is also 0 for any $K : A \times S \to \mathbb{R}$. Therefore, any η satisfying the constraint for some λ must be 0 for all (a, s), so DIA is satisfied.

For necessity, if DIA holds then the value of the dual is always 0 for any $K : A \times S \to \mathbb{R}$. By strong duality, the value of the primal is also 0 for any K. Therefore, given K, there is a feasible primal solution $\xi_i(a, s)$ that satisfies all primal constraints, and CEB holds. □

Theorem 7. Fix an arbitrary action profile $\widehat{a} \in A$ and consider the following disobedient deviation plan α_i for every agent i: always play $\widehat{a}_i$ regardless of the mediator's recommendation a_i and report s_i with probability $\Pr(s_i|a_i, \widehat{a}_{-i}) = \sum_{s_{-i}} \Pr(s|a_i, \widehat{a}_{-i})$ independently of the actual signal realization. If any agent i unilaterally deviates according to α_i, the probability of reported signals becomes

$$\Pr(s|a, \alpha_i) = \begin{cases} \Pr(s_1|\widehat{a}) \Pr(s_2|\widehat{a}) & \text{if } a_1 = \widehat{a}_1 \text{ and } a_2 = \widehat{a}_2 \\ \Pr(s_1|\widehat{a}) \Pr(s_2|\widehat{a}_1, a_2) & \text{if } a_1 = \widehat{a}_1 \text{ and } a_2 \neq \widehat{a}_2 \\ \Pr(s_1|a_1, \widehat{a}_2) \Pr(s_2|\widehat{a}) & \text{if } a_1 \neq \widehat{a}_1 \text{ and } a_2 = \widehat{a}_2 \\ \Pr(s_1|a_1, \widehat{a}_2) \Pr(s_2|\widehat{a}_1, a_2) & \text{if } a_1 \neq \widehat{a}_1 \text{ and } a_2 \neq \widehat{a}_2 \end{cases}$$

These probabilities are the same regardless of who deviates, hence IOA fails. □

Theorem 8. Given the ordering of agents in the main text, if $k > 0$ permute agent k with agent 1 and consider the following block matrix (blank spaces denote blocks of zeros).

$$Q = \begin{bmatrix} Q_1 & Q_1 & Q_1 & Q_1 & Q_1 \\ -Q_2 & & & & \\ & -Q_3 & & & \\ & & \cdots & -Q_{n-1} & \\ & & & & -Q_n \end{bmatrix}$$

where Q_i is the matrix with $(|A_i|\,|S_i|)^2$ rows and $|A|\,|S|$ columns defined pointwise by

$$Q_i(a_i, s_i, b_i, t_i)(\widehat{a}, \widehat{s}) = \begin{cases} \Pr(\widehat{s}_{-i}, t_i|\widehat{a}_{-i}, b_i) & \text{if } (a_i, s_i) = (\widehat{a}_i, \widehat{s}_i) \\ 0 & \text{otherwise.} \end{cases}$$

Now, IOA is satisfied if

$$\lambda Q = 0 \text{ and } \lambda \geq 0 \quad \Rightarrow \quad \lambda_i(a_i, s_i, b_i, t_i) = 0 \text{ whenever } a_i \neq b_i. \tag{$*$}$$

To see this, by definition IOA holds if $\lambda_i(a_i, b_i, \rho_i) = 0$ for all $a_i \neq b_i$ whenever $\lambda \geq 0$ and $\eta \in \mathbb{R}^{A \times S}$ satisfy $\sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) = \Lambda$ for all (i, a_i) and

$$\forall(i, a, s), \qquad \sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) \Pr(s|a_{-i}, b_i, \rho_i) \;=\; \eta(a, s).$$

Adding these equations with respect to s for all (i, a) yields $\sum_s \eta(a, s) = \Lambda$, so we may drop the constraints $\sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) = \Lambda$. Rearranging, the left-hand side above becomes

$$\sum_{(b_i,\rho_i)} \lambda_i(a_i, b_i, \rho_i) \sum_{t_i \in \rho_i^{-1}(s_i)} \Pr(s_{-i}, t_i|a_{-i}, b_i) \;=\; \sum_{(b_i,t_i)} \sum_{\{\rho_i : \rho_i(t_i) = s_i\}} \lambda_i(a_i, b_i, \rho_i) \Pr(s_{-i}, t_i|a_{-i}, b_i).$$

Write $\lambda_i(a_i, s_i, b_i, t_i) = \sum_{\{\rho_i : \rho_i(t_i) = s_i\}} \lambda_i(a_i, b_i, \rho_i)$. Now, IOA holds if $\lambda_i(a_i, s_i, b_i, t_i) = 0$ whenever $a_i \neq b_i$ for any $\lambda \geq 0$ and η such that

$$\forall(i, a, s), \qquad \sum_{(b_i,t_i)} \lambda_i(a_i, s_i, b_i, t_i) \Pr(s_{-i}, t_i|a_{-i}, b_i) \;=\; \eta(a, s),$$

from which $(*)$ now follows. Let $\hat{Q}$ be the matrix derived from Q by removing the following redundant rows and columns. One row of Q is redundant because for every agent $i > 1$,

$$\forall (\widehat{a}, \widehat{s}), \qquad \sum_{(a_1,s_1)} Q_1(a_1, s_1, a_1, s_1)(\widehat{a}, \widehat{s}) \;=\; \sum_{(a_i,s_i)} Q_i(a_i, s_i, a_i, s_i)(\widehat{a}, \widehat{s}).$$

There may also be redundant column vectors. If $k > 1$, fix any agent $i \leq k$ with $i > 1$ and any $(a_1, a_i) \in A_1 \times A_i$. Then, for any $\widehat{a}$ such that $\widehat{a}_1 = a_1$ and $\widehat{a}_i = a_i$,

$$\forall (b_1, b_i), \qquad \sum_{\widehat{s}} Q_1(a_1, b_1)(\widehat{a}, \widehat{s}) \;=\; 1 \qquad \text{and} \qquad \sum_{\widehat{s}} Q_i(a_i, b_i)(\widehat{a}, \widehat{s}) \;=\; 1.$$

Therefore there are $|A_{-1,i}| - 1$ redundant columns for each (a_1, a_i). There are even more redundant column vectors when $k > 1$. Fix any $(a_1, a_i) \in A_1 \times A_i$ for $i \leq k$. Note that there exist $\widehat{a}$ such that $\widehat{a}_1 = a_1$ and $\widehat{a}_i = a_i$ for which no column has been deleted as redundant (indeed there exists only one such $\widehat{a}$ due to the previous step). Denote such $\widehat{a}$ by $\widehat{a}(a_1, a_i)$. Now, for any $a_1' \neq a_1 \in A_1, a_i' \neq a_i \in A_i$, we have

$$\begin{aligned}
\forall (b_1, b_i), \qquad \sum_{\widehat{s}} Q_1(\widetilde{a}_1, b_1)(\widehat{a}(a_1', a_i'), \widehat{s}) \;&=\; \sum_{\widehat{s}} Q_1(\widetilde{a}_1, b_1)(\widehat{a}(a_1', a_i), \widehat{s}) \\
&\quad + \sum_{\widehat{s}} Q_1(\widetilde{a}_1, b_1)(\widehat{a}(a_1, a_i'), \widehat{s}) - \sum_{\widehat{s}} Q_1(\widetilde{a}_1, b_1)(\widehat{a}(a_1, a_i), \widehat{s}) \\
\sum_{\widehat{s}} Q_i(\widetilde{a}_i, b_i)(\widehat{a}(a_1', a_i'), \widehat{s}) \;&=\; \sum_{\widehat{s}} Q_i(\widetilde{a}_i, b_i)(\widehat{a}(a_1', a_i), \widehat{s}) \\
&\quad + \sum_{\widehat{s}} Q_i(\widetilde{a}_i, b_i)(\widehat{a}(a_1, a_i'), \widehat{s}) - \sum_{\widehat{s}} Q_i(\widetilde{a}_i, b_i)(\widehat{a}(a_1, a_i), \widehat{s})
\end{aligned}$$

for $\widetilde{a}_1 = a_1', a_1$ and $\widetilde{a}_i = a_i', a_i$ (all other rows are 0 and these equations are trivially satisfied). Hence there are $(|A_1| - 1)(|A_i| - 1)$ more redundant columns for $(1, i)$. If $|S_{-n}| = 1$ a similar argument shows that there are $|A_n|\,|S_n|\,(|A_n| - 1)$ additional redundant rows.

By construction, IOA holds if $\lambda \hat{Q} = 0$ implies that $\lambda = 0$. This holds generically if (1) $\hat{Q}$ has full row rank generically, i.e., it has no more rows than columns, so

$$\begin{aligned}
\sum_{i=1}^{n} (|A_i|\,|S_i|)^2 - 1 - \chi_n\,|A_n|\,|S_n|\,(|A_n| - 1) \;&\leq\; (n-1)\,|A|\,|S| - \sum_{i=2}^{k} |A_1|\,|A_i|\,(|A_{-1,i}| - 1) \\
&\qquad - \sum_{i=2}^{k} (|A_1| - 1)\,(|A_i| - 1) \\
&=\; (n-1)\,|A|\,|S| - (k-1)(|A| - |A_1| + 1) + \sum_{i=2}^{k} |A_i|\,,
\end{aligned}$$

where $\chi_n = 1$ if $|S_{-n}| = 1$ and 0 otherwise, and (2) each $\hat{Q}_i$ has full row rank generically. If $|S_i| > 1$ and $|S_{-i}| > 1$ then (2) is implied by $|A_i|\,|S_i| \leq |A_{-i}|\,|S_{-i}|$. If $|S_i| = 1$ then (2) is implied by $|A_i| - 1 \leq |A_{-i}|\,(|S_{-i}| - 1)$ after removing redundant columns as a result of $\sum_{\widehat{s}} \widehat{Q}_i(a_i, b_i)(\widehat{a}, \widehat{s}) = 1$ for all $\widehat{a}$. Finally, the case $|S_{-i}| = 1$ was treated in the previous paragraph. This completes the proof. □

Theorem 9. We use the following notation. Given a correlated strategy σ and a deviation plan α_i, let $\Delta v_i(\sigma,\alpha_i) = \sum_{(a,b_i,\rho_i)} \sigma(a)\alpha_i(b_i,\rho_i|a_i)(v_i(a_{-i},b_i)-v_i(a))$ be the utility gain from α_i at σ and $\Delta \Pr(s|a,\alpha_i) = \sum_{(a,b_i,\rho_i)} \alpha_i(b_i,\rho_i|a_i)(\Pr(s|a_{-i},b_i,\rho_i) - \Pr(s|a))$ the change in the probability that s is reported from α_i at a. Enforcing an arbitrary correlated strategy σ subject to participation constraints reduces to finding transfers ζ to solve the following family of linear inequalities:

$$\begin{aligned} \forall(i,a_i,b_i,\rho_i),\ \sum_{a_{-i}} \sigma(a)(v_i(a_{-i},b_i)-v_i(a)) \ &\leq \sum_{(a_{-i},s)} \sigma(a)\zeta_i(a,s)(\Pr(s|a_{-i},b_i,\rho_i) - \Pr(s|a)), \\ \forall(a,s), \qquad \sum_{i=1}^{n} \zeta_i(a,s) \ &= \ 0, \\ \forall i \in I, \qquad \sum_{a\in A} \sigma(a)v_i(a) - \sum_{(a,s)} \sigma(a)\zeta_i(a,s)\Pr(s|a) \ &\geq \ 0. \end{aligned}$$

The dual of this problem subject to participation is:

$$\max_{\lambda,\pi\geq 0,\eta} \sum_{i\in I} \Delta v_i(\sigma,\lambda_i) - \pi_i v_i(\sigma) \ \text{ s.t. } \forall(i,a,s), \quad \sigma(a)\Delta\Pr(s|a,\lambda_i) \ = \ \eta(a,s) + \pi_i\sigma\Pr(s|a)$$

where π_i is a multiplier for agent i's participation constraint and $v_i(\sigma) = \sum_a \sigma(a)v_i(a)$. Adding the dual constraints with respect to $s \in S$, it follows that $\pi_i = \pi$ does not depend on i. Redefining $\eta(a,s)$ as $\eta(a,s)+\pi\Pr(s|a)$, the set of feasible $\lambda \geq 0$ is the same as without participation constraints. Since $\sum_i v_i(a) \geq 0$ for all a, the dual is maximized by $\pi = 0$. □

Theorem 10. We use the same notation as in the proof of Theorem 9. Let $z = (z_1,\ldots,z_n)$ be a vector of liability limits for each agent. Enforcing σ subject to limited liability reduces to finding ζ such that

$$\begin{aligned} \forall(i,a_i,b_i,\rho_i), \quad \sum_{a_{-i}} \sigma(a)(v_i(a_{-i},b_i)-v_i(a)) \ &\leq \sum_{(a_{-i},s)} \sigma(a)\zeta_i(a,s)(\Pr(s|a_{-i},b_i,\rho_i) - \Pr(s|a)), \\ \forall(a,s), \qquad \sum_{i=1}^{n} \zeta_i(a,s) \ &= \ 0, \\ \forall(i,a,s), \qquad \zeta_i(a,s) \ &\leq \ z_i. \end{aligned}$$

The dual of this metering problem subject to one-sided limited liability is given by:

$$\max_{\lambda,\beta\geq 0,\eta} \sum_{i\in I} \Delta v_i(\sigma,\lambda_i) - \sum_{(i,a,s)} \beta_i(a,s)z_i \ \text{ s.t. } \forall(i,a,s),\ \sigma(a)\Delta\Pr(s|a,\lambda_i) = \eta(a,s) + \beta_i(a,s),$$

where $\beta_i(a,s)$ is a multiplier on the liability constraint for agent i at (a,s). Adding the dual equations with respect to s implies $-\sum_s \beta_i(a,s) = \sum_s \eta(a,s)$ for all (i,a). Therefore,

$$-\sum_{(i,s)} \beta_i(a,s)z_i \ = \ \sum_{(i,s)} \eta(a,s)z_i \ = \ \widehat{z}\sum_{s\in S} \eta(a,s),$$

where $\widehat{z} = \sum_i z_i$, so we may eliminate $\beta_i(a,s)$ from the dual and get the equivalent problem:

$$\max_{\lambda \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \lambda_i) + \widehat{z} \sum_{(a,s)} \eta(a,s) \quad \text{s.t. } \forall (i,a,s),\ \sigma(a) \Delta \Pr(s|a,\lambda_i) \ \geq \ \eta(a,s).$$

Any two liability profiles z and z' with $\widehat{z} = \widehat{z}'$ lead to this dual with the same value. □

Theorem 11. We use the same notation as in the proof of Theorem 9. Enforcing σ subject to participation and liability is equivalent to the value of the following problem being zero:

$$\min_{\zeta} \sum_{(i,a_i)} \varepsilon_i(a_i) \quad \text{s.t.} \quad \forall (i,a,s),\ \zeta_i(a,s) \ \leq \ z_i, \qquad \forall (i, a_i, b_i, \rho_i),$$

$$\sum_{a_{-i}} \sigma(a)(v_i(a_{-i}, b_i) - v_i(a)) \ \leq \ \sum_{(a_{-i},s)} \sigma(a) \zeta_i(a,s)(\Pr(s|a_{-i}, b_i, \rho_i) - \Pr(s|a)) + \varepsilon_i(a_i),$$

$$\forall (a,s), \qquad \sum_{i \in I} \zeta_i(a,s) \ = \ 0,$$

$$\forall i \in I, \qquad \sum_{a \in A} \sigma(a) v_i(a) - \sum_{(a,s)} \sigma(a) \zeta_i(a,s) \Pr(s|a) \ \geq \ 0.$$

The first family of constraints imposes incentive compatibility, the second budget balance, the third individual rationality, and the last corresponds to one-sided limited liability. The dual of this metering problem is given by the following program, where λ, η, π and β represent the respective multipliers on each of the primal constraints.

$$\max_{\alpha, \pi, \beta \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I} \pi_i v_i(\sigma) - \sum_{(i,a,s)} \beta_i(a,s) z_i \quad \text{s.t. } \forall (i, a_i), \sum_{(b_i, \rho_i)} \alpha_i(b_i, \rho_i | a_i) \ = \ 1$$

$$\forall (i,a,s), \qquad \sigma(a) \Delta \Pr(s|a, \alpha_i) \ = \ \eta(a,s) + \pi_i \sigma(a) \Pr(s|a) + \beta_i(a,s).$$

Adding the dual constraints with respect to $s \in S$, it follows that

$$-\sum_{(a,s)} \beta_i(a,s) = \sum_{(a,s)} \eta(a,s) + \pi_i \ = \ \widehat{\eta} + \pi_i$$

where $\widehat{\eta} := \sum_{(a,s)} \eta(a,s)$. After substituting and eliminating β, the dual is equivalent to

$$V \ := \ \max_{\alpha, \pi \geq 0, \eta} \sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \widehat{z} \quad \text{s.t.}$$

$$\forall (i,a,s), \qquad \sigma(a) \Delta \Pr(s|a, \alpha_i) \ \geq \ \eta(a,s) + \pi_i \sigma(a) \Pr(s|a).$$

Now, σ is enforceable if and only if $V = 0$, i.e., if and only if for any dual-feasible (α, π, η) such that $\sum_i \Delta v_i(\sigma, \alpha_i) > 0$, we have that

$$\sum_{i \in I} \Delta v_i(\sigma, \alpha_i) \ \leq \ \sum_{i \in I} \pi_i (v_i(\sigma) - z_i) + \widehat{\eta} \widehat{z}.$$

Finally, since the dual objective is increasing in η, an optimal solution for η must solve

$$\eta(a,s) \ = \ \min_{i \in I} \{ \Delta \Pr(s|a, \alpha_i) - \pi_i \Pr(s|a) \}.$$

This completes the proof. □

Corollary 1. Given the dual problem from the proof of Theorem 11, the first statement follows because if $v_i(\sigma) \geq z_i$ then the objective function is decreasing in π_i and reducing π_i relaxes the dual constraints. The second statement follows by rewriting the objective as

$$\sum_{i \in I} \Delta v_i(\sigma, \alpha_i) - \sum_{i \in I \setminus t} \pi_i(v_i(\sigma) - z_i) + \widehat{\eta} \sum_{i \in I} z_i,$$

where t is the set of agents whose participation constraint won't bind ($\pi_i^* = 0$ for $i \in t$). □

References

Alchian, A. and H. Demsetz (1972): "Production, Information Costs, and Economic Organization," *American Economic Review*, 62, 777–795. 1

Aoyagi, M. (2005): "Collusion Through Mediated Communication in Repeated Games with Imperfect Private Monitoring," *Economic Theory*, 25, 455–475. 15

Aumann, R. (1974): "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, 1, 67–96. 1

d'Aspremont, C., J. Cremer, and L.-A. Gérard-Varet (2004): "Balanced Bayesian Mechanisms," *Journal of Economic Theory*, 115, 385–396. 12

d'Aspremont, C. and L.-A. Gérard-Varet (1998): "Linear Inequality Methods to Enforce Partnerships under Uncertainty: An Overview," *Games and Economic Behavior*, 25, 311–336. 1, 14, 15, 16

Fudenberg, D., D. Levine, and E. Maskin (1994): "The Folk Theorem with Imperfect Public Information," *Econometrica*, 62, 997–1039. 2, 16

Holmström, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340. 1, 3, 11

Kandori, M. (2003): "Randomization, Communication, and Efficiency in Repeated Games with Imperfect Public Monitoring," *Econometrica*, 71, 345–353. 2, 15

Kosenok, G. and S. Severinov (2008): "Individually Rational, Balanced-Budget Bayesian Mechanisms and the Allocation of Surplus," *Journal of Economic Theory*, 140, 126–261. 15

Legros, P. and H. Matsushima (1991): "Efficiency in Partnerships," *Journal of Economic Theory*, 55, 296–322. 1, 14, 15

Legros, P. and S. Matthews (1993): "Efficient and Nearly Efficient Partnerships," *Review of Economic Studies*, 60, 599–611. 1, 2, 14, 15

Myerson, R. (1986): "Multistage Games with Communication," *Econometrica*, 54, 323–358. 1

Radner, R., R. Myerson, and E. Maskin (1986): "An Example of a Repeated Partnership Game with Discounting and with Uniformly Inefficient Equilibria," *Review of Economic Studies*, 53, 59–69. 1, 3, 4

Rahman, D. (2008): "But Who Will Monitor the Monitor?" Mimeo. 11, 12

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press. 17

Tomala, T. (2005): "On Subgame-Perfect Communication Equilibria in Repeated Games with Imperfect Monitoring," Mimeo. 15